

696
1954

THE LIBRARY OF CONGRESS CLASSIFICATION: CONTENT ANALYSIS OF THE SCHEDULES

IN PREPARATION FOR
THEIR CONVERSION INTO
MACHINE-READABLE FORM

The Library of Congress Classification:
A Content Analysis of the Schedules
in Preparation for their Conversion into
Machine-Readable Form

WITHDRAWN
FROM THE VPI & SU
LIBRARY COLLECTION

Deposited by U.S. Gov't

JUL 12 1995

VPI & S.U. Library

Nancy J. Williamson
Principal Investigator

With the Assistance of
Suliang Feng

and

Tracy Tennant

Library of Congress Cataloging-in-Publication Data

Williamson, Nancy Joyce.

The Library of Congress classification : a content analysis of the schedules in preparation for their conversion into machine-readable form / Nancy J. Williamson ; with the assistance of Suliang Feng and Tracy Tennant.

p. cm.

Includes bibliographical references.

ISBN 0-8444-0871-9

---- Copy 3 Z663.72.L492 1995

1. Classification, Library of Congress--Data processing. I. Feng, Suliang. II. Tennant, Tracy. III. Title.

696.U4W53 1995

025.4'33'0285--dc20

94-49342

CIP



Copyright® 1995 the Library of Congress, except within the U.S.A.

ISBN 0-8444-0871-9

For sale by

The Library of Congress, Cataloging Distribution Service
Washington D.C. 20541
(202) 707-6100

TABLE OF CONTENTS

	<u>Page Nos.</u>
1. PREFACE	i
2. ACKNOWLEDGEMENTS	iii
3. EXECUTIVE SUMMARY	v
4. INTRODUCTION	1
5. PREVIOUS RESEARCH	2
6. THE STUDY	4
6.1 Basic Objectives	6
6.2 Methodology	7
7. CONTENT ANALYSIS OF THE LCC SCHEDULES	10
7.1 General Characteristics of the Schedules	10
7.1.1 Recommendations	16
7.2 Hierarchical Arrangement	17
7.2.1 Recommendations	28
7.3 Tables and Subarrangements	28
7.3.1 General Characteristics	28
7.3.2 Subarrangements	33
7.3.3 Tables	35
7.3.3.1 Screen Displays of Large Tables	38
7.3.3.2 Manipulation of Tables	40
7.3.3.3 Summary of Findings	43
7.3.4 Recommendations	45
7.4 Notes and References	46
7.4.1 General Characteristics	46
7.4.2 See References	49
7.4.2.1 Independent See References	51
7.4.2.2 See References Which Relocate Topics	53
7.4.2.3 "For ... see ..." References	55
7.4.2.4 Problems With References	56

7.4.3 Notes	61
7.4.3.1 "Cf. ..." Notes	62
7.4.3.2 "Prefer ..." Notes	71
7.4.3.3 "Class here" and "Including notes"	72
7.4.4 Recommendations	75
7.5. Footnotes	76
7.5.1. Recommendations	79
7.6 Class Numbers	79
7.7 Divided Like Instructions	80
7.7.1 Recommendations.	82
7.8 Caption and Record Size	82
7.8.1 Classification Notation	82
7.8.2 Captions	83
7.8.3 Notes and "For ... see..." References	84
7.8.4 Other Classification Data Elements	85
8. SUMMARY OF RECOMMENDATIONS	86
9. FURTHER RESEARCH	88
10. CONCLUSION	91
11. APPENDICES	94
A. Schedules Used in the Study	95
B. Schedule Sampling Form	98
C. Instructions for Sampling the LCC Schedules	100
D. References	103

1. PREFACE

It was a fortuitous event for both the Library of Congress (LC) and Nancy J. Williamson when during the International Federation of Library Associations and Institutions (IFLA) conference in August 1987, Nancy Williamson and Lucia Rather, then the Director for Cataloging, met and discussed activities at LC concerning the Library of Congress Classification (LCC) System. During 1987, LC's Processing Services Department (now Collections Services) had formulated functional requirements for the automation of the Library of Congress Classification (LCC) and outlined plans for action. LC's goal was to develop an online system to enable the Library to 1) maintain the classification schedules; 2) perform shelf listing; 3) use classification to enhance subject access; and 4) produce from the file the schedules in a variety of formats, e.g. hard copy, CD-ROM. In addition, the benefits to be derived by other libraries were also taken into consideration.

All of the above called for the development of a machine-readable format, both as the basis for use internally and also for distribution. It was recognized early on that it was necessary to know the characteristics of LCC before proceeding with the detailed format design.

During this same period of time, Nancy Williamson had requested and obtained permission for a sabbatical leave from the Faculty of Library and Information Science, University of Toronto, to start in January 1988. Based on findings of other researchers in classification systems, reported in the literature, and on her own analysis of LCC, she had determined that her sabbatical would be spent on classification research. To date, LCC had received less attention than its two counterparts, the Universal Decimal Classification System (UDC) and the Dewey Decimal Classification System (DDC). Thus the chance meeting in 1987 provided the foundation required for both parties to proceed, with a research project leading to the development of a machine-readable format and eventually an automated system for LCC. The research project was announced by LC in January 1988 at the ALA midwinter meeting.

This report contains the results of that study: a detailed content analysis of the Library of Congress Classification Schedules. It provides an overview of the system for which individual schedules were completed separately, often by different classifiers, over a long period of time. In doing so, it identifies the various similarities and differences among the schedules. The data in the study was used by the Library of Congress in its design of the MARC Classification Format, which was given provisional approval by the ALA MARBI Committee in June 1990. It is also influencing the conversion of the schedules to machine-readable form. Over the past two years, the Library of Congress has converted almost half the LC classification into the MARC format, and the full set of schedules in machine-readable form are targeted to appear in 1995.

The Library commends principal investigator Nancy J. Williamson for a job well done and thanks the organizations which supported her work. Her efforts should enhance the development of the next phase of improving subject retrieval using the capabilities of the Library of Congress Classification System.

Sarah E. Thomas
Director for Cataloging

2. ACKNOWLEDGEMENTS

In the course of this research, numerous individuals and organizations have provided advice, assistance and financial support without which the research could not have been accomplished. Most importantly, the staff of the Library of Congress gave encouragement, technical assistance and advice throughout the course of the project. In particular, my sincere thanks go to Henriette Avram, former Associate Librarian for Collections Services and to Lucia Rather, former Director for Cataloging, who were instrumental in initiating the project and who gave support and advice during the planning stages. Also, the principal investigator is indebted to Mary Lou Miller, Senior Automation Planning Specialist, Automation Planning & Liaison Office, who was principal consultant throughout the formulation of the study and the data gathering. Her expertise and combined knowledge of the MARC formats and the Library of Congress Classification were invaluable to the research. My thanks and appreciation must also go to Mary Kay Pietris, Cataloging Policy and Support Office, Julianne Beall, Assistant Editor, Dewey Decimal Classification, and Rebecca Guenther, Senior MARC Specialist, Network Development & MARC Standards Office for their interest and assistance at various stages of the project.

Funding for the project was absolutely essential if the research was to be carried out. I am most grateful and deeply indebted to the following organizations for their generous support of the project:

Canadian Library Association
Faculty of Library and Information Science, University of Toronto
Humanities and Social Sciences Research Committee, University of
Toronto
Gale Research Company
Online Computer Library Center (OCLC Inc.)
International Federation of Library Associations and Institutions (IFLA)
Standing Committee on Classification and Indexing

Five of the above organizations provided research grants totalling \$22,550.00 (Can.) and the Gale Research Company donated the Library of Congress Classification Schedules (valued at \$6000.00 Can.). As well as

providing some funds through its Centre for Research in Library and Information Science, the Faculty of Library and Information Science, University of Toronto provided office space, computer support and secretarial assistance.

My sincere thanks also to my two research assistants both of whom were students at the Faculty of Library and Information Science. Suliang Feng, senior research assistant, carried out the major portion of the data collection, designed the database, using dBase IV, recorded all of the data in machine-readable form and carried out the statistical analysis. He was ably assisted in the data collection by Tracy Tennant.

Finally, my grateful thanks and appreciation to Marcia Chen, Faculty Secretary at the Faculty of Library and Information Science, who so patiently and carefully formatted the text and designed the figures contained in this report.

Nancy J. Williamson,
Professor and Principal Investigator
Faculty of Library and Information Science
University of Toronto

3. EXECUTIVE SUMMARY

This report summarizes the findings from a content analysis of the data elements contained in the printed version of the Library of Congress Classification schedules. The purposes of the analysis were to identify characteristics of schedule data that have implications for the editing and conversion of the schedules into machine-readable form and to make suggestions and recommendations that might aid those responsible for making the decisions necessary to carry out these processes.

The characteristics of seven types of data were analysed, including the general characteristics of the hierarchical display of topics and subtopics, the tables and Cutter subarrangements, notes and references, footnotes, the number of classification numbers, divided like instructions and caption size. Two kinds of recommendations have been made, those which are absolutely essential at the beginning stages of the conversion and those which it is strongly recommended be considered for the future. Specific recommendations for each aspect of the analysis are enumerated at the end of each Section of the report. Based on the findings in this research it is recommended that:

1. The integrity of the LCC schedules be maintained but the use of new technological devices be used wherever they could improve the manipulation of the schedules;
2. All data which are only implicit on the printed page be made explicit in the machine-readable version;
3. Errors and peculiarities in the hierarchical relationships among captions be corrected so that the correct relationships are maintained in the displays of the machine-readable records;
4. The display requirements for schedules and tables needed for various users be determined - USMARC record display, page display, hierarchical and other displays;

5. Where clarification is needed, data elements be edited and clarified for effective search and retrieval;
6. Consideration be given to editing some data elements, such as notes and cross references, after the schedules are in machine-readable form, rather than at the input stage, so that advantage may be taken of machine search capability during the editing process;
7. Duplicate tables and subarrangements be consolidated wherever possible, and the potential use of a common set of Cutter numbers across schedules be considered;
8. The use of an expert system for the manipulation of tables and subarrangements be investigated;
9. The simultaneous display of schedule and table data for the use of classifiers be investigated;
10. Notes and references be made more explicit and all cross references to class numbers by accompanied by meaningful captions;
11. Divided like instructions be eliminated and the pertinent class numbers and their captions inserted into the appropriate locations in the schedules;
12. A study of the indexes in the individual schedules be undertaken to determine how they may be most efficiently and effectively integrated into the system;
13. Priorities be set for the editing requirements. Which are essential, useful and desirable, and non-essential; at what point in the conversion process should they be carried out?
14. If the printed schedules are to continue to be produced, the relationship between the machine-readable and printed versions of the schedules be determined.

For the future it is further recommended that:

1. A study be undertaken to determine the feasibility of eliminating all, or most, of the tables in the LCC system and the means by which this might be accomplished; and
2. The possibility of setting up policies and guidelines for the editors of the LCC system be explored to ensure that the same kinds data are more consistent across the LCC system (e.g., the function and wording of notes and references and guidelines for the creation of index terms).

The Library of Congress Classification system is very heterogeneous indeed. For every example taken from the data exceptions can probably be found. In an analysis of this kind it would be virtually impossible identify absolutely every peculiarity of the schedules. However, it is the hope of the investigators that the sample of 3934 pages examined, out of 12,267 pages of schedules is sufficient to identify the major problems which the LCC will present in its conversion to a machine-readable system. This analysis has indeed demonstrated that there are principles and patterns which apply across much of the data. The conversion of the LCC schedules to machine-readable form will be a mammoth task because of the tremendous amount of editing that appears to be required. Nevertheless, with clear goals, an understanding of users' needs, sound planning, patience, time and considerable effort it should be possible to create an effective machine-readable product. If the conversion can be accomplished effectively, the result should be extremely useful to LCC editors, classifiers and other users.

The nature of this project has resulted in a very lengthy process. The Library of Congress has been kept informed of major findings as the work progressed. A completed draft of the report was submitted to LC in September 1993 and the principal investigator made a presentation to LC staff in December 1993. By that time LC was already working on the conversion process and had implemented some of the recommendations of this report.

4. INTRODUCTION

The Library of Congress Classification (LCC) is a very large and complex system. Because of its heterogeneous nature, the computerization of the LCC schedules might at first seem to be an impossible task. Of course it is not. As is the case with other numeric and alphabetic data, the LCC schedules are indeed convertible to machine-readable form. However, if the end-product is to be successful, there are important questions for which it is essential to have authoritative answers. Can the LCC schedule data be stored and manipulated so that they can be accessed intelligently, usefully displayed and efficiently manipulated for effective use in a computerized environment? If so, what would conversion to machine-readable form entail and how should it be accomplished? What special requirements are there for use of the schedules online? What kinds of editing and how much will be required as part of the conversion? What improvements could be incorporated into the system, so that LCC could be a better managed classification than it is in the printed mode? What should be the long-term goals of the project which, if completed, could have far reaching effects on the LC cataloguing staff, subject cataloguers in individual libraries, public service librarians and catalogue users? What would be the sheer volume of the work to be accomplished? These are some of the broad questions addressed in this research.

There is considerable precedent for serious investigation into the ways and means of developing a system for the use of LCC online. Other classification systems have been converted to machine-readable form for practical use. However, the general characteristics of LCC are such that a more detailed analysis of the schedules than has previously existed is needed as the basis for answering a number of fundamental questions.

It was the primary task of the investigators to document the nature and characteristics of the LCC schedules as they presently exist and to identify features of the system which could affect decisions to be made with respect to

their conversion to machine-readable form and to consider possible implications for data manipulation and use in a new physical environment. Both quantitative and qualitative data were collected and analyzed. Particular attention was given to the amount of data to be converted, to discernible patterns in the data and to the discovery of inconsistencies, idiosyncrasies, anomalies and technical errors.

It was not the purpose of this study to analyze the philosophical, intellectual and structural strengths and weaknesses of LCC. Nor was it one of the goals to suggest changes in the intellectual content or a restructuring of the LCC schedules. Rather the objective was to categorize and describe the data as they appear in the schedules, and insofar as possible to make suggestions and recommendations as to possible solutions to some of the problems.

5. PREVIOUS RESEARCH

Sophisticated manipulation of book classification systems to enhance computerized retrieval is not a new idea. It can be traced back more than 45 years to numerous research projects which are well documented in the literature of library and information science (Rigby 1974, 1981). Many of these projects were launched between 1948 and 1980. Not surprisingly, individual research projects were to some extent isolated and extremely varied in nature. Experimentation ranged from the maintenance and display of schedules, through the indexing of bibliographic files, vocabulary control and thesaurus construction, to the display of bibliographic records and the use of classification in online retrieval. Not unnaturally, the Universal Decimal Classification (UDC) was prominently featured in this early research. Its notation could easily be manipulated by computer and to a considerable extent it reflects the hierarchical relationships among its classes and subclasses. Of particular interest in terms of classification research is the development and experimentation with an online interactive retrieval system called "AUDACIOUS" in which Robert Freeman and Pauline Atherton Cochrane

(Freeman and Atherton 1968) investigated the use of the UDC schedules for browsing and for organizing bibliographic files in the nuclear sciences.

The technology used in the early research would, today, be considered primitive at best and almost all of the work predated the existence of online catalogues. However, these pioneering efforts were an important beginning and an invaluable lead-in to classification research in online catalogues in the 1980's. Moreover, times have changed. Computer technology has become much more flexible and much more sophisticated, while, over the past 13 years, the concern for improving the quality of subject access in online catalogues has intensified. New approaches to the manipulation and use of classification online is one aspect of this concern. As with UDC, the Dewey Decimal Classification (DDC) has a notation which is easily manipulated by computer and which in many instances reflects the hierarchical relationships between topics and their subtopics, making it a logical choice for research and development in the 1980's. Two important projects resulted - the development of the Dewey Decimal Classification Editorial Support System (Beall 1992) and the Dewey Decimal Classification Online Project (Markey and Demeyer 1986). Obviously there are major philosophical and structural differences between DDC and LCC. Nevertheless, when it comes to categorizing and describing their data elements, there are many similarities. Both have class numbers, topics or captions, cross references, hierarchical structure, notes and instructions. Whatever the classification system, together these projects provided a strong foundation for new developments in the maintenance and use of classification systems online.

There may be many reasons why early consideration was not given to a machine-readable LCC. Undoubtedly, size of the schedules must have been an important factor. Certainly the prospect of converting more than 40 volumes of some 12,300 pages of schedules (exclusive of outlines and indexes) to machine-readable form could be somewhat daunting. Also the lack of a hierarchical notation would surely have given cause for hesitation.

Nevertheless, the DDC Online Project has demonstrated the value of classification as a tool in online retrieval and Cochrane and Markey (1985) predicted "that some day there will be a machine-readable version of the Library of Congress Classification". New and more sophisticated ways of handling data have emerged and an analysis of prospects for LCC by Lois Chan (1986) and Nancy Williamson (1986, 1989) suggested that such a goal might be attainable.

Such a project presents enormous challenges, but it is an essential part of a much more complex long-term goal in the total computerization of storage and retrieval operations. Currently, serious research is being carried out to develop a cataloguers' workstation which could ultimately permit the entire library cataloguing operation to take place at a computer terminal. Already in existence are machine-readable bibliographic records, subject and name authority control systems in machine-readable form, and local shelflists accessible through online systems. As well, the computerization of the Anglo-American Cataloguing Rules 2nd edition, 1988 revision is under discussion. The facility to use LCC online could be the final "jewel in the crown" of computerized catalogues and catalogue support systems of the future. Clearly, very careful planning and considerable time and effort will be involved, but the potential is there for this to be accomplished in the foreseeable future.

6. THE STUDY

For a number of reasons the very nature of LCC made it desirable to carry out a content analysis of the schedules in preparation for their actual conversion to machine-readable form. Prior to this study there was insufficient documentation on LCC in existence at the appropriate level of detail needed to support all of the decisions required for the development of a machine-readable format and the editing of the schedules for conversion. LCC is massive in size and its historical development is such that it is a collection of systems, rather

than a single coherent well integrated scheme. It is true that there are general principles which apply across the system, but each class together with its subclasses has been developed separately based on literary warrant and the existence of published literature. Each discipline has its own unique features and tables, and there are few mnemonics applied universally throughout the system. Moreover most, but not all, schedules have their own detailed indexes, but there is no overall official LC index to the whole classification.

Particularly important from the perspective of computerization is the LCC notation which is not hierarchical in nature. Hierarchical relationships are inherent only in the relative display of captions. To preserve the systematic relationships among captions, special handling will be required in the machine environment. LCC is also very page-oriented. Data, such as class numbers, are usually only partially complete and are only understood within the context of the page display. Furthermore although Immroth's Guide to the Library of Congress Classification (Chan 1990) is an invaluable support tool, there are few instructions and interpretations embedded in the schedules themselves to aid in the use of LCC and its tables. A content analysis would clarify uncertainties and reveal problems to be solved.

From the outset of the study, it was necessary to establish a context, or frame of reference, within which to work. This context was provided through general objectives, primarily based on the perceptions of the Library of Congress staff and the principal investigator, of the possibilities and requirements for the future uses of LCC in online systems. Reality dictated that the development of a computerized system, or systems, for the maintenance and use of LCC would be evolutionary and that the basic machine-readable format would need to be capable of supporting all of the various objectives.

6.1 Basic Objectives

The investigation developed from two basic assumptions. A MARC or MARC-like record format would be used to accommodate the data elements and provide for their manipulation and display to serve a variety of functions. Also it was assumed that the ultimate goal would be the design and development of an integrated system which might, over a period of time, enable the Library of Congress to achieve four clearly defined objectives:

- 1) The design and implementation of an online, interactive support system to enable the Library of Congress to edit, update and maintain its classification schedules in an online system;
- 2) The provision of a system which would permit classifiers at the Library of Congress, and in other libraries, to classify and shelflist online;
- 3) The production of the LCC schedules in a variety of physical forms, including printed volumes, computer tapes, and CD-ROM; and
- 4) The design of a catalogue users' search tool which would enhance subject retrieval in online catalogues.

In keeping with the possible development of a cataloguers' workstation, the classification schedules might ultimately become part of a complex system which would require links among classification schedules, bibliographic records, authority control systems and shelflists.

In a broader context, it was the hope that the machine-readable format could be adapted internationally to accommodate other classification systems such as the Dewey Decimal Classification, the Universal Decimal Classification and the National Library of Medicine Classification. Indeed as the study progressed, the proposal for a USMARC format has been adapted to accommodate DDC, while editors and users of other classification systems have been looking seriously at the possibilities of the MARC format for classification for their own use.

A system of this magnitude and complexity must necessarily evolve one function at a time and be based on careful research and planning from the outset. The development of the editorial support system is an essential first step in the successful achievement of the remaining three objectives. Thus the machine-readable record format would need to provide for all data, both schedule and non-schedule, required to permit the various kinds of output and display needed to support all of the proposed functions.

6.2 Methodology

Since no previous study of this kind existed, careful planning, procedures and rules for the data gathering, and analysis were necessary. The project progressed through 4 distinct phases - a planning phase, an exploratory phase, a pre-test, and a data gathering and analysis phase.

Planning began in the Fall of 1987 when the principal investigator met with the Library of Congress staff. The objectives and some directions for the project were established and in January 1988, during the ALA Midwinter Meeting in San Antonio, the project for the computerization of LCC was formally announced. As the first step in the planning process a prototype MARC record for classification was prepared at LC. Long experience with the conversion of data to machine-readable form by LC staff had led to the conclusion that the most expedient approach to the research was to begin with a record format, however tentative that format might be. For purposes of the research this proved to be a very astute decision. This prototype format, best described as a working format for purposes of research, paralleled the USMARC format for authority records and combined in a single record all of the data types identified from the LCC schedules, tables and indexes. The prototype assumed that each caption, independent see reference and node label, together with its accompanying data, would become a separate record. Records would be derived from both schedules and tables.

This prototype record served several functions essential to the progress of the study. It provided a focus for the content analysis of the schedules and played an essential role in establishing definitions and in identifying and categorizing the more unique features of LCC. Data elements from the schedules were matched against the prototype which was revised and updated as the research progressed. The prototype format ultimately evolved into a format acceptable to LC, and was submitted to the American library community for discussion through the American Library Association's MARBI (Machine-Readable Bibliographic Information) Committee. In June 1990, with modifications, this format was approved by the USMARC Advisory Group as a provisional format for classification. Subsequently the provisional format was published by LC's Cataloging Distribution Service as the USMARC Format for Classification Data: Including Guidelines for Content Designation (Library of Congress. Network Development and MARC Standards Office 1990).

The second phase, the exploratory phase, served several purposes. It was a period in the research for testing the prototype format, identifying peculiarities of the data not hitherto discovered, and determining the questions to be answered in the data gathering process. During this phase, the principal investigator and Mary Lou Miller of the LC staff applied the format to a spectrum of the LCC schedules which might be expected to present problems. Initially, what appeared to be new types of data were located in every few pages examined. However as this preliminary analysis of the data progressed, there were fewer occurrences of new data types and definite patterns in the data emerged. As a result of the exploration of the schedules, eight potential questions relating to schedule data were drawn up. With supporting documentation, these questions were submitted to LC's statistician for advice on sampling the data. Seven of the eight questions were selected for use in the data gathering. A random sample was recommended, using 100 pages per schedule for large schedules; 50 pages for schedules of less than 100 pages. Finally a set of Instructions for Sampling (Appendix C) were drawn up and a "Schedule Sampling Form" (Appendix B) for recording the data was prepared.

The third phase of the project, the pre-test, took place over a period of 3 months in mid-1988. Its purpose was to test the sampling form and to identify problems with the sampling techniques. The pre-test resulted in minor adjustments to the both the form and the instructions for sampling and identified types of qualitative data which needed to be collected.

Proceeding with the data gathering and analysis phase 3,934 pages or 32% of the pages across 42 schedules were sampled. All data on each page were examined and described on the sampling form under the following categories:

- 1) The hierarchical display of topics or captions;
- 2) The nature and characteristics of the tables and subarrangements;
- 3) References and notes;
- 4) Footnotes;
- 5) The number of single class numbers, spans of numbers and independent see references;
- 6) The "divided like" instructions and the classification numbers encompassed by those instructions; and
- 7) The size of the captions.

LCC schedules selected for use in the study were the most complete schedules in one sequence available at the time that the final content analysis began. This included 38 schedules from the Gale Research Company's publication Library of Congress Classification Schedules Combined with Additions and Changes Through 1986 (Library of Congress 1986) and the latest available editions of G, H-HJ, R and Z from LC's own schedules (Appendix A). This combination was used because a complete set of the Gale 1986 edition was not available at the beginning of the project. All references in the report to specific schedules relate to the editions used in the study. As the data gathering was completed validation procedures were carried out by the principal investigator and, for purposes of analysis the data were entered into a

computer database using dBASE IV. A by-product of the validation was the collection of additional qualitative data which identified unique features of individual schedules. One of these by-products was the cross-checking of "see" and "For ... see ..." references and cf. notes to determine accuracy and consistency.

7. CONTENT ANALYSIS OF THE LCC SCHEDULES

7.1 General Characteristics of the Schedules

A typical LCC schedule contains a class or one or more subclasses and may have up to 5 distinct sections. Schedules with more than one subclass usually contain a "synopsis" or list of subclasses contained in the volume. In most schedules there is also an "outline" which provides an overview of the topics in each subclass. The outline is followed by the detailed schedules, major tables, and, where appropriate, an index. Not all schedules have either a "synopsis" or an "outline" and these terms are used inconsistently across the system. For example in schedule P-Pa, "Synopsis I" is correctly named, while "Synopsis II" has the characteristics of an "outline". While not analyzed in this study, these synopses and outlines could be even more important in access to LCC online than they are in the printed volumes. As with any kind of information system an overview equivalent to a "table of contents" can provide classifiers and other users with an essential frame of reference for orientation to the LCC system. Serious consideration needs to be given to the possible role which synopses and outlines, in some form, might play in the computerized environment. The outlines have not been developed consistently across the system but perhaps edited versions of the outlines could become a feature of an interface between users and the detailed schedules.

In the printed version of LCC, the indexes are an essential feature of individual schedules. Some schedules have very detailed indexes. Others, notably some subclasses in Class P, had no indexes at all. While a detailed analysis of the indexes has not been part of this study, there is little doubt that

these indexes vary in both terminology and depth of indexing across the LCC system. The variations occur as a result of the nature of the schedules and LC's policy of revising one schedule at a time. Also, as with other aspects of the system, the indexes have been created individually and the depth of indexing is dependent on the subject matter and the perspective of the indexer. Nevertheless, the indexes are extremely rich in terminology which, when included in the MARC records, should enhance the online access to the machine-readable schedules. However, it is important to recognize that these indexes have been created for individual schedules, over time, without benefit of a general indexing policy for the whole of the LCC system. The precise effect of the existing indexes on access and retrieval in this new environment is unknown. However, the indexing requirements for online systems differ somewhat from those of printed systems. In their current state the integration of the individual indexes into the database could both help and hinder online retrieval. An analysis of the indexes was not part of this study and further investigation is needed. Degree of coverage, consistency and specificity of terminology, and the ability to counteract scatter of subjects by disciplines are among the requirements for enhancing access to the schedules.

While there are slight differences in the displays used in the LC and Gale editions of the schedules, they are minor and would not be a factor in the conversion of the schedules. Neither set of schedules uses special typographical devices in its displays. However, the nature of some disciplines requires the use of non-Roman alphabets.

CLASS Q		
QH	BIOLOGY (GENERAL)	QH
475	Reproduction - Continued Asexual Cf. QH442.2 Cloning (Genetic engineering)	
477	Fission	
479	Budding	
481	Sexual	
485	Fertilization	
487	Parthenogenesis	
489	Alternation of generations	
491	Development. Morphogenesis Cf. QK665, Plant embryology QL951-991, Animal embryology	
499	Regeneration Cf. QK840, Plant physiology QP90.2, Animal physiology	
	Life Cf. QK710-899, Plants QP1-348, Animals	
501	General works, treatises, and textbooks	
	Biochemistry, <u>see</u> QH345	
504	Biomagnetism Cf. QP82.2.M3, Physiological effects of magnetic fields QP345, Human magnetic fields	
505	Biophysics	

Figure 1. Schedule Format

As can be seen in Figure 1 taken from the Gale (1986) schedules, each page begins with the class letter followed by the letters and name of the subclass in upper case letters. Captions which are the topics and subtopics are set out in systematic order with each subtopic being indented three spaces to the right indicating subordination and hierarchical relationship. Independent "see" references, as for example "Biochemistry, see QH345", and Cutter numbers with topics, such as the list at "TL798 Special types [of Artificial satellites], A-Z" in Figure 2, are part of the hierarchical display.

CLASS T

TL	MOTOR VEHICLES. AERONAUTICS. ASTRONAUTICS.	TL
796.5	Astronautics. Space Travel Artificial satellites By country, region, etc., A-Z - Continued e.g. United States .U5 General works .U6 Special projects or programs, A-Z e.g. .U6V3 Vanguard	
.6	By planet, etc., A-Z e.g. .E2 Earth .S8 Sun	
.8	Observers' handbooks	
797	Space stations	
798	Special types, A-Z .G4 Geodetic satellites Cf. QB343, Artificial satellites in geodesy .M4 Meteorological satellites Cf. QC879.5, Meteorology etc.	

Figure 2

Notes, examples and dependent "for... see..." references are further indented under the captions to which they belong. Many pages contain small tables, which provide for subarrangements using successive Cutter numbers. Other, larger tables may take one or several full pages. Some of these are embedded in the schedules; other are at the ends of the volumes. Each page of schedules contains the complete hierarchical context for the topics and subtopics beginning with the subclass. To ensure this, parts of the hierarchy which carry over from a previous page are repeated at the top of the page, the last repeated caption being followed by the word "continued". Some pages have footnotes. Most often these are instructions referring the classifier to actual page numbers elsewhere in the schedule or to a particular table.

Typical footnotes are the following taken from Class N - "For Table II see pp. 345-353. Add country number in table to 4000," and from Class V - "For subarrangement, see tables under schools, p. 14." In a few instances, footnotes are actually small tables in the form of Cutter subarrangements. Examples of

the latter can be found in Subclass QL (Gale 1986).

Terminology of the captions may be distinctive, as for example "Biophysics", "Nutrition" and "Respiration", or non-distinctive, as for example "General works", "Special parts", "Special topics, A-Z", and "Other". Both types must be viewed in the context of the hierarchies to which they belong. Non-distinctive captions require the context to even be meaningful at all. Both distinctive and non-distinctive captions require the context to ensure that the aspect of the subject being treated is seen in the context of the class or subclass to which it belongs.

As can be seen in Figures 1 and 2, many of the classification numbers are incomplete and can be understood only in the context of the page on which they appear. The subclass letters appear only at the top of the page and provide the context for the whole page. Class numbers linked with specific captions begin with the numerical part of the notation and appear in the left margin. Cutter lists show only the Cutter number and depend on the page layout to show the class number and caption to which they belong. Some complete class numbers may be found in notes but these are usually class numbers from other parts of the schedules. Even here some abbreviation of numbers is used. For example in "Cf. QP1-348, Animals" under "Life" in Figure 1 the QP has not been repeated before 348. Moreover, some captions display no notation at all, although their position in the array of topics and subtopics must be preserved. Captions unaccompanied by class numbers are of two types - independent "see" references, and captions which function as node labels for a number of subtopics. For example in Figure 1, preceding QH501 the caption "Life" has no class number of its own. General discussions of this topic would be classified in "QH501 General works, treatises and text books". However, "Life" serves as a node label and encompasses the subtopics assigned notations ranging from QH501 through QH531. Implicit in the display is "QH501-QH531 Life". In a computer-based system the position of both node labels and independent "see" references must be preserved in the systematic

arrangement and the spans of numbers represented by node labels must be made explicit.

The page display and the indentation of topics is not always consistent. Sometimes individual subtopics are indented more than the three spaces to the right of their superordinate topics. In some cases it is evident that this is a result of a transcription error in, for example, the Gale edition of the schedules; in others it may have been done for reasons of display. For example in Figure 2 at "TL798 Special types, A-Z", the caption for TL798.G4, "Geodetic satellites", is indented 9 spaces to the right of the indentation for "Special types". In reality, this caption is the next level in the hierarchy below "Special types" and would be correctly placed three spaces in (under the "e" of "Special"). This is typical of many such displays throughout the schedules. The relationship is easily understood when scanning the printed page. However, in a computerized system the page environment will no longer aid in the interpretation of the context, so it will be important in the conversion to a MARC format to maintain the precise relationships among topics and subtopics if they are to be correctly displayed on a computer screen. Tables also present display problems and these are discussed in Section 7.3.3 on "Tables".

From the sample pages examined across the LCC system, it was readily apparent that, as presented in printed form, the LCC data are very page-dependent. This is not necessarily a bad thing in printed schedules. Too much data on a printed page can be just as much of a problem as too little. At the same time, some enhancement of the data might improve the printed version. All print-dependent data and incomplete data will need to be edited for the machine-readable version of LCC. However, if it is assumed that there will continue to be a printed product and that it will be produced from the USMARC records, it will be important to give some thought to any differing display requirements for the two versions of the schedules. Should some of the data which must be made explicit for the computer remain implicit in the printed version? If so, which data and why? What effect, if any, will the

different outputs have on the USMARC record? It can also be concluded from the content analysis that, while major portions of the data are recognizable and easily categorized, there are numerous instances throughout the system which require intellectual decisions to determine the exact nature of the data and methods for handling them.

7.1.1 RECOMMENDATIONS

Based on the findings on the general characteristics of the schedules, it is recommended that:

- a) The integrity of the schedules in printed form be maintained, but also that the machine-readable version be designed to respond to the advantages of the computer environment;
- b) The role of synopses and outlines in the computerized environment be determined;
- c) All implicit and page dependent data be made explicit for the machine-readable version and rules be established which will make the conversion as speedy, accurate and efficient as possible;
- d) Where feasible and appropriate, non-distinctive captions be enhanced to make them more effective and more easily comprehensible. (For example, under "Nuclear fission", "Special topics, A-Z" might be extended to read "Special topics in nuclear fission, A-Z");
- e) Care be taken during the conversion process to ensure that intellectual decisions are applied when necessary;
- f) Any differences in the USMARC record requirements for the various products be determined; and
- g) A separate study be carried out on the existing individual indexes to determine their nature, characteristics, and the most effective means of utilizing them in the USMARC records. In the longer term a general indexing policy is needed which will meet the requirements for a general index to the computerized LCC schedules.

7.2 Hierarchical Arrangement

Hierarchical relationships are the essence of all classification. Enumerative classification systems provide a systematic arrangement of subjects according to a set of principles based on an accepted philosophy of the organization of knowledge, on patterns established on the basis of literary warrant and, frequently, on a combination of both. However, classified order is not self-evident. Some method or device is required to preserve the relationships among classes, subclasses, topics and subtopics. Depending on the system, this may be accomplished in different ways. In some classification systems, for example DDC, these relationships are preserved and may be manipulated through a hierarchical notation. LCC does not fit this pattern. Its notation preserves order but it does not reflect hierarchy. Rather LCC's hierarchical relationships are preserved in the display of captions on the printed page and when the data are removed from the printed page and stored and manipulated in a computer system, some other means must be found to preserve those relationships.

The selection of an effective method for presenting the systematic display is essential to the conversion of the data to machine-readable form and the ultimate use of the schedules. There are at least 3 possible approaches which might be taken. One method would be to superimpose a numbering system which would be invisible to the user but which would preserve the hierarchical relationships and permit automatic renumbering as the classification is maintained and updated. A second approach might be to link records hierarchically. However, the linking process could be very complex and maintenance and updating of the schedules could be both time consuming and error prone and might affect the speed with which the system could reconstruct the hierarchies for purposes of display. A third method, the one chosen to be incorporated into the provisional USMARC, is to embed the hierarchy that provides the context for each caption in the record for that caption. Since this method repeats data across records it is highly redundant, but it has other

advantages. For example, it should be relatively simple to edit, maintain and update the schedules and it should be helpful for the schedule editors to be able to see all of the data related to a particular caption displayed together in a single record. Based on present information, this latter method appears to be the most expedient solution. Of course, there may be other solutions not yet thought of. It will be essential in the testing of the system to determine whether the provisional USMARC record format satisfies the requirements for maintenance and updating as well as for use by classifiers. At least 4 types of schedule display are worth considering:

- a) A USMARC record display;
- b) A browsable hierarchical display of each caption and its subtopics in its hierarchical context (Figure 3);
- c) A browsable linear display similar to the one now seen in the printed schedules (Figure 4); and
- d) A step by step or "menu-driven" approach, combined with an "explode" or "expand" command (Figures 5-7).

S AGRICULTURE		
SF	ANIMAL CULTURE	SF
SF277-SF360.3	Horses	
SF321-SF359.7.M3	Racing	
SF351-SF357.58	Running races. Flat racing	
SF357.5-SF357.58	Quarter racing	
SF357.5	General works	
SF357.53	History	
SF357.55	By region or country, A-Z	
SF357.57-SF357.575	Biography, <u>see</u> SF336	
	Horses	
	Cf. SF293.Q3, Quarter horse breed	
SF357.57	Collective	
SF357.575	Individual, A-Z	
SF357.58	Individual races, A-Z	

Figure 3 Hierarchical Display

While the kind of screen display ultimately provided for in the system may differ from that shown above, the enhanced display in Figure 3 exemplifies the information which classifiers and other users should be able to see if "Quarter racing" were selected for viewing in a hierarchical mode. The display should provide the context for "Quarter racing" and it should also be possible to display topics which are subordinate to "Quarter racing". These characteristics of display will be important in both classifying and updating the schedules.

A linear browsable display would provide a different approach and enable users to scroll through the schedules viewing each caption as it follows another, as shown in Figure 4. This approach may be of importance to classifiers who have been in the habit of scanning whole pages of the printed schedules and still feel that this is a useful approach. Also there are occasions when there is an advantage to being able to view a topic in relation to its neighbours. This is not always possible in a strictly hierarchical display.

S AGRICULTURE

SF	ANIMAL CULTURE	SF
	... etc. {11 and 1/2 pages}	
SF277-SF360.3	Horses	
	For anatomy, <u>see</u> SF765	
	For physiology, <u>see</u> SF768.2.H67	
SF277	Periodicals. Societies. Serials	
SF321-SF359.7.M3	... etc. {5 and 3/4 pages}	
SF321	Racing	
	Cf. HV6718, Racetrack gambling	
	Periodicals. Serials	
	Cf. SF323, Clubs and societies	
SF321.2	Congresses	
SF321.5	Dictionaries. Encyclopedias	
	Including terminology, racing	
	slang, etc.	
SF323.A1A-Z-SF323.A3-Z	Clubs and societies	
	Including reports, history,	
	programs, etc.	
SF323.A1A-Z	General works	
SF323.A3-Z	Individual clubs and societies.	
	By name, A-Z	
SF324-SF324.4	Racetracks	
SF324	General works	
SF324.3-SF324.4	By region or country	
SF324.3-SF324.35	United States	
SF324.3	General works	
SF324.35.A-SF324.35.Z	By region or state	
SF324.4.A-SF324.4.Z	Other regions or countries	
SF351-SF357.58	... etc. {1 page}	
	Running races. Flat racing ... etc.	

Figure 4 Linear Display

However the schedules are voluminous and linear scrolling may be cumbersome and time consuming at best. As indicated in Figure 4 particular topics can be many pages from the beginning of a class or subclass.

For example, in the S schedule there are eleven and one-half pages from the beginning of "SF Animal Culture" to the location of its subordinate topic "Horses". "Racing" which is subordinate to "Horses" is an additional five and three-quarters pages from the beginning of "Horses". While this kind of scrolling permits browsing, it may not be the most effective method of searching. In the long term this kind of scrolling may become outmoded as a

method of searching the schedules. However it may be the most comfortable approach for current users.

A fourth type of display might be a "menu" approach which would facilitate a step-by-step move down the hierarchy beginning with the captions contained in the USMARC record display as illustrated in Figures 5 through 7.

STEP 1: Find "Racing"

S AGRICULTURE		
SF	ANIMAL CULTURE	SF
SF277-SF360.3	Horses	
SF321-SF359.7.M3	Racing	

Figure 5

STEP 2: Expand "Racing"

S AGRICULTURE		
SF	ANIMAL CULTURE	SF
SF277-SF360.3	Horses	
SF321-SF359.M3	Racing	
SF351-SF357.58	Running races. Flat racing	
SF359-SF359.7.M3	Steeplechase and hurdle racing	

Figure 6

STEP 3: Expand "Running races. Flat racing"

S AGRICULTURE		
SF	ANIMAL CULTURE	SF
SF277-SF360.3	Horses	
SF321-SF359.7.M3	Racing	
SF351-SF357.58	Running races. Flat racing General works, <u>see</u> SF334 History, <u>see</u> SF335 Betting, <u>see</u> SF331+ Training of runners Individual stables, A-Z Individual running horses, A-Z Individual races, A-Z Quarter racing	
SF351		
SF353		
SF355		
SF357		
SF357.5-SF357.58		

Figure 7

In all cases, it is essential to display each caption in the context of its complete hierarchy, but Figures 3 through 7 illustrate the fact that data will be displayed differently depending on which method, or methods, of display are provided for in the computerized system. Further study will be needed to determine which approaches to viewing the data are useful and to design displays which work to users' advantage.

In addition to the display of captions, information was gathered on the complexity of the hierarchies and the problems that they might present in displaying data. Would the redundancy of embedding the whole hierarchy in every record be a serious problem? Would the hierarchies be so cumbersome that they would hamper the screen display? Do the schedules contain many extremely long hierarchies?

It was assumed that every caption, including the subclass, every independent see reference and every caption in a Cutter list is part of the hierarchies of topics and subtopics. Examples, attached to Cutter numbers such as those illustrated in Figure 2 were also assumed to be topics in the

hierarchies. Each page in the sample was examined and, as indicated by section 1 of the "Schedule Sampling Form" in Appendix B, the number of captions at each level in the hierarchies was recorded. The class or subclass named at the top of each sample page was designated as level 1. Captions at the extreme left of each table display were also level 1 captions. Where errors or anomalies in the hierarchies were detected these were "corrected" in the process of counting the captions. The resulting data were used to calculate and to analyze the nature of the hierarchies found in the schedules and tables. This served to answer such questions as: How many captions were located at level "n"? How long are the hierarchies? How many captions are there at a particular hierarchical level?

The hierarchical levels of 132,213 captions from the sample were recorded. The longest hierarchies located had 17 levels and 4 such hierarchies were found under the individual works of Goethe in Subclass PT, Part 1, German Literature (Gale 1986). However, these hierarchies were very much the extreme. The largest number of captions, 27,603 or 20.8% of the captions, were located at level 5. The next largest groups were 25,379 captions at level 4 and 24,483 at level 6. 77,465 or 58.5% of the captions fell within levels 4-6. Moreover the number of captions at any level fell off rapidly above level 7. For example, there were 1,585 captions at level 11 and 507 at level 12. From the data gathered, it is reasonable to conclude that the large majority of LCC hierarchies are not so excessively long and complex that they could not be easily accommodated on a computer screen.

Finally, in converting the schedule and table data to machine-readable form, it is absolutely essential that the correct hierarchical relationships be maintained. It is also essential to recognize that there are instances in which the visual presentation of the captions in the printed schedules can be misleading with respect to their precise hierarchical relationship with other captions. Without adjustment data could be "lost" in the system. In collecting the data on the hierarchical levels of the captions, 5 potential problems were

identified. These are as follows:

- a) Some minor typing errors were found in the hierarchies. The nature of the printed schedules is such that visually it is not always easy to be absolutely certain of the indentations. Some of these errors appeared to be the result of misinterpretation or miscalculation in preparing the Gale Research schedules. Also, some examples were found where in moving to a new page one step in the hierarchy has been left out. It often takes only one mistake to misplace the rest of a hierarchy.
- b) Some data, particularly the captions in the Cutter lists and examples with class numbers, are not precisely lined up at the correct indentation in their hierarchies and their relationship with other captions may need to be determined on an ad hoc basis. These are not errors in the usual sense but probably result from efforts to make the page format either clearer or more attractive. Figure 2 illustrates some examples of this problem.
- c) In some parts of the schedules, the topic immediately subordinate to the subclass changes in the middle of the page. These new topics look very like the captions of the subclasses, since they are centred on the page and in upper case letters. However, in reality they are subordinate to the subclass and any other upper case centred heading which appears at the head of a page. Most mid-page upper case centred headings are at level 2, although in Class J some cases were found where a mid-page heading was really at level 4. In Subclass JN the hierarchy "Constitutional History and Administration - Europe - Italy - United Italy (1870-)" is an example of this, where "United Italy (1870-)" is at hierarchical level 4 and the 3 superordinate captions are all centred headings in upper case letters (Gale, 1986; Class J., p. 227). The result is that all of the captions subordinate to a mid-page centred heading move down one or more levels even though the position on the page may suggest otherwise; and

d) There is one practice which has been built into the schedules over the years and which could potentially cause problems in scanning list of names, institutions, etc. in the machine-readable version of the schedules. A straight alphabetical list of names is easily browsable and it would be expected that all the names would appear at one hierarchical level. However, several instances were found in the schedules where the names are not all displayed at one level. In such cases there are spans of letters (e.g. N - NZ) or spans of names as exemplified in Figure 8 (e.g. Binet - Caro) which act as node labels for groups of names, while other names, have entries of their own and are not subsumed under node labels. For example, a node label may be at level 8 and the names gathered under that node label at level 9, while some individual names in the list are at level 8. Examples of this kind of display can found in PG Russian Literature "Individual authors' works, 1800-1870", in the list of philosophers in B-BJ and in lists of museums in Class N. Figure 8 illustrates the problem.

B-BJ PHILOSOPHY. PSYCHOLOGY		
B	PHILOSOPHY (GENERAL)	B
B790-B5739	Modern (1450/1600-)	
B850-B5739	By region or country	
B1801-B2430.W47	France	
B1815-B2430.W47	By period	
B2185-B2417	19th century	
B2189-B2417	Individual philosophers 1/	
B2189	A - Alliot	
B2190	Alliot, François	
B2190.5	Alliot - Ampère	
B2191	Ampère, André Marie	
B2192	Ampère - Binet	
B2192.B3	Bautain, Louis Eugène Marie	
B2192.B4	Bernard, Claude	
B2193	Binet, Alfred	
B2194	Binet - Caro	
B2194.B4	Bonald, Louis Gabriel Ambroise, viscomte de	
B2194.B5	Boutroux, Émile	
B2195	Caro, Elme Marie	
B2197	Caro - Comte	
B2200-B2249	Comté, Auguste (1) etc.	

Figure 8

In Figure 8 the captions under "Individual philosophers" are of two kinds. Those captions which contain dashes, for example "Ampère - Binet", are node labels which assign a particular class number to a group of names which would then be subarranged by Cutter numbers. As indicated by the example, "Ampère - Binet" will include all the names after Ampère, André Marie and before Binet, Alfred. Other names, such as "Binet, Alfred" have their own class number or, as exemplified by "Comté, Auguste", their own span of numbers. As a result, two different kinds of data, names and node labels, are at the same hierarchical level and individual names are at two different levels. The distinction appears to be related to LCC's pragmatic use of space. In this case the distinction seems to have been made between major philosophers which have their own class numbers or spans of numbers, and minor philosophers which have been handled using Cutter numbers. Also, it can be assumed that the LC shelflist may contain additional particular names which have been assigned the node label class numbers together with their unique Cutters. From this and other examples, it appears that the practice of using the node labels has been partially discontinued. For instance in subclasses B-BJ, the French "Individual philosophers" of the 20th century are listed in one alphabetical sequence with no node labels and all names at the same level in the hierarchy. Nevertheless, in the same schedule the practice of using node labels has been continued under both England and Germany. This could be because the "Individual philosophers" of the 19th and 20th centuries have been grouped together and 20th century philosophers interfiled with a previously existing 19th century list. Other lists using this practice, which are located in other parts of the schedules have slight variations but the general problem of node labels which affect the hierarchies is the same.

Two approaches to this problem appear possible. One solution would be to subsum all names under node labels; the other would be to remove the node labels from the schedules. Both solutions would result in a list of names all of which would appear at the same level in a particular hierarchy. Does this kind of node label serve a useful purpose? If it does, this is not apparent from

analyzing the schedules and it would appear to add complexity to both the maintenance and use of the schedules. What problems would arise if the node labels were removed? This question would need to be addressed by comparing the instances of this phenomenon with LC's own shelflist. However, in sampling the schedules, there appeared to be no obstacles to the removal of this particular type of node label. In all cases examined the classification numbers could be applied without the node labels being present. However, LC might want to consider making the lists complete by adding to the schedules any class numbers and names from its shelflist which do not now appear in the schedules.

The most effective browsable display for a user would be a straight list of names without having to manipulate the node labels. Also a straight list of names should be easier to maintain and update. More investigation may be needed to determine the most effective method of handling this problem and consideration should be given to a consistent approach across schedules in future.

e) Finally there is sometimes a question as to what are schedule data and what are table data. On some pages of schedules which contain tables, the relationship between the parts of the schedules which precede and follow a table can be confusing. In some cases it is important to consider such questions as: Does this caption belong to the schedules or to the table? If it belongs to the schedules, at what level in the hierarchy does it belong?

The situations described above may, or may not, be the only instances of technical problems in the hierarchical display of the schedules and tables but they are sufficient to illustrate the need to be aware of problems of this kind as well as the need to exert care in the conversion process.

7.2.1 RECOMMENDATIONS

It is recommended that:

- a) In the conversion to machine-readable form care be taken to ensure that the correct hierarchical relationships are maintained and errors corrected as necessary; and
- b) Where node labels have been used in lists of names, a means be found whereby the lists can be displayed as a continuous list without the interspersion of node labels.

7.3 *Tables and Subarrangements*

7.3.1 *General Characteristics*

LCC's tables might be described as the nerve centre which drives the classification system. Together with the "divided like" mechanisms, tables are devices which make possible the development of a very precise enumerative system, while permitting economy of space in the printed schedules. At the same time, the tables add complexity to the system, its manipulation and display. The majority of LCC tables are tailored to individual schedules rather than to the classification system as a whole. Therefore, not surprisingly, there are numerous variations from one table to another across the system.

Nevertheless, findings show that there are some general principles which apply and some repetition of patterns in table data across schedules. There are large tables and small tables, simple tables and complex tables, tables within tables and alternative tables. Also the procedures for their manipulation vary. Some tables such as those in Classes J and K are unique to those particular classes.

For purposes of the study, both the "internal" tables embedded in the schedules and the "external" tables located at the end of subclasses and/or physical volumes were analyzed. As indicated by Section 2 of the "Schedule Sampling Form" in Appendix B, tables were divided into two major types, based on function - tables and subarrangements. The term "table" was used to

describe tables which provide for identification of topics and subtopics and are applied to spans of numbers. The term "subarrangement" was used to categorize tables which subarrange topics under single class numbers using Cutter numbers, or further subarrange topics themselves represented by Cutter numbers. For example, in Figure 9, the tables of "10 nos.", "5 nos." and "2 nos." were counted as "tables", while the table for "1 no." was counted as a "subarrangement". Small internal tables which provide for successive Cutting fall into the category of "subarrangements".

SUBARRANGEMENTS			
HV SOCIAL PATHOLOGY. SOCIAL AND PUBLIC WELFARE. CRIMINOLOGY HV			
HV697-HV4959	Protection, assistance and relief		
HV697-HV1493	Special classes		
HV701-HV1420.5	Children		
HV959-HV1420.5	Orphanages. Orphans		
HV971-HV1420.5	By region or country - Continued		
HV1001-HV1420.5	Other regions or countries. Table IX 1/ <u>{LEVEL 1}</u>		
<u>10 nos.</u>	<u>5 nos.</u>	<u>2 nos.</u>	<u>1 no.</u>
(1)	(1)		.A1-5 Collected works (nonserial)
(4)	(2)	(1)	.A6-Z7 General works
			Including history
(5)			Early through 1800
(6)			1801-
(8)	(3)		Other
		(2)	Local, A-Z
(9)	(4)		By region or state, A-Z
(10)	(5)		By city, A-Z
			Under each state or city:
			.x General works
			.x2 Particular
			institutions, A-Z
<u>{LEVEL 2}</u>		<u>{LEVEL 3}</u>	
1/ For Table IX, <u>see</u> pp. 195-204. Add country number of table to 1000.			

Figure 9. Three Levels of Tables

In hindsight this distinction appears to be somewhat nebulous. However, to some extent, it does separate tables into those tables which require simple manipulation and those which require more complex procedures for application.

Data collected concerning tables and subarrangements included:

- a) The number of tables and/or subarrangements on a page;
- b) The types of data included - topical, form and/or geographical;
- c) The "level" or distance of the table or subarrangement from the main schedule; and
- d) The types of notation used to arrange topics - classification numbers, Cutter numbers, or both.

The sample of 3,934 pages identified 220 pages or 5.5% of the sample which contained table data. In addition there were 556 subarrangements most of which were small tables providing for successive Cutting. A concentration of tables and subarrangements appear in Classes C, H-HJ and HM-HX.

Classes K and P are special cases. Schedule P-PZ Literature Tables contained 44 tables for use across Class P but no schedules. This reduced the number of tables which appeared in the other volumes of class P. A search of these latter schedules located 31 tables on 3.6% of the 853 Class P pages in the sample. Based on these figures it is estimated that there would be a potential for 90 small tables in Class P, in addition to the 44 large external tables in the P-PZ Literature Tables.

Class K is quite different from the other schedules. Some schedules, for example, KF Law of the United States and KK-KKC Law of Germany contained considerably more pages of tables than pages of schedules. There were 1,846 pages of schedules and tables in the seven "K" schedules in the sample (Gale 1986). Of these pages 615 pages or approximately 33% of the pages in the sample from the K volumes were large tables exclusive of small subarrangements. When the statistics for all of the schedules were analyzed it was estimated that this particular edition of LCC could potentially contain 1,171 pages of tables, in addition to 1,729 subarrangements. Since very large tables may cover a number of pages this does not mean that there would be 1,171 large tables. Rather it means that approximately 9% of the pages in the 41 schedules could contain table data. While in quantity this does not

represent a large percentage of the total pages, the tables are intense and frequently several tables are combined with common captions (Figure 9). Their conversion to machine-readable form will require a great deal of care. Moreover they are unevenly spread across the LCC system and some schedules will be much more difficult to convert than others.

In describing the types of data, topic and form were grouped together because most tables and subarrangements of this kind contained both, whereas some tables and subarrangements include only geographical data. In other instances all three types of data were present. With respect to the term "level" some tables and subarrangements are themselves in a hierarchy-like relationship to each other. Thus, with respect to tables, the term "level" describes the order in which the tables and subarrangements are applied to the schedules in particular cases where more than one table and/or subarrangement is required to develop a particular class number. In the data collection, the question answered was "Is this the 1st, 2nd or 3rd table used in determining a class number?

Figure 9 illustrates this feature. An instruction may lead the classifier to a table at the end of the schedule (LEVEL 1) then back to a table internal to the schedule (LEVEL 2) and the internal table may itself require further subarrangement using successive Cutters (LEVEL 3) in order to determine the final class number. This "hierarchy" of tables controls citation order within the topics. To take a particular instance, a classifier might wish to determine the classification number for the "Assistance and Protection of orphans in Toronto, Canada". The Classifier would first locate HV1001-HV1420.5, the span of numbers to be applied to the "Protection, assistance and relief of orphans" in "Other regions or countries". From this point the classifier is directed to "Table IX" (LEVEL 1) to locate the information that Canada is represented by numbers 1-10 in this sequence (i.e. HV1001-HV1010). From the table for "10 nos." (LEVEL 2) in the schedules, it is determined that the 10th number (i.e. HV1010) represents cities and is further subarranged "By city, A-Z". Here the

classifier is led to a table for subarranging by successive Cutters (LEVEL 3) based on whether or not the item in hand is a "General work" or about a "Particular institution". For a "General work" on "Assistance and Protection of Orphans in Toronto, Canada" the resulting classification number would be HV1010.T66.

The primary function of both tables and subarrangements is to provide a method for extending the classification system without significantly increasing the bulk of the schedules. Tables and subarrangements have a number of characteristics in common. Both:

- a) Have data types similar to the main schedules;
- b) Contain hierarchical displays of topics and subtopics accompanied by cross references and notes;
- c) In relation to the schedules, are subordinate to the schedules and are linked either to single class numbers or spans of class numbers in the schedules;
- d) Have for the most part been developed to be used with particular classes, subclasses or individual class numbers rather than to be applied across a number of schedules, or the whole classification system; and
- e) Include a notation, of some kind. The kind of notation varies. These notations are not "class numbers" in the same sense as the classification notation used in the main schedules. In some cases the notation may be part of a class number from the schedules to which they apply; in other cases they may be surrogates for class numbers and Cutter numbers. The kind of notation determines the kind of manipulation required. Among the types of notations found in tables are:
 - i) Surrogates for Cutter numbers, for example .x and .x2;
 - ii) Lists of Cutter numbers;

- iii) Arabic numbers which refer to the "n"th number in the sequence, for example (1), (2), etc.; and
- iv) The final digits, or pairs of digits contained in a classification notation sequence, for example 0, 1, 2, 3, 4, or 5, 6, 7, 8, 9.

In both tables and subarrangements there is some duplication across schedules and there are several structural patterns that have been applied across the system. The major differences between tables and subarrangements are those of size, complexity and function.

7.3.2 Subarrangements

Subarrangements are the simplest of the two types to describe and manipulate. For purposes of data-gathering, "subarrangements" were defined as formal structures, separate from the schedules, which extend the topics and/or subtopics under a topic represented by a single class number or Cutter number. Within the limits of this definition 4 distinct types of subarrangements were located in the sample.

The most prevalent type of subarrangement is the small table or subarrangement, usually only two or three captions in length, which instructs the classifier to assign successive Cutter numbers. These subarrangements are unnumbered, are usually introduced by the instruction "under each" and are embedded in the schedules under the particular class number to which they are to be applied. For example at HD1491.A5-Z the schedule caption is "By region or country, A-Z" followed by the instruction:

Under each country:
 .x General works
 .x2 Local, A-Z

For this particular type, the identical captions and notations occur under other

topics at HD6331.2 and at HD6338.2 on the same page of schedules. There is much duplication of this same subarrangement throughout the schedules wherever there is a caption "By region or country, A-Z". In some sections of the classification these small subarrangements exist in profusion. In Class L, for example, there were instances of this same subarrangement appearing 4 times and 5 times on the same printed page (Gale 1986; Class L, p. 101). The same pattern is prevalent across schedules. For example in subclass "TL Motor Vehicles. Aeronautics. Astronautics" in Figure 10 similar data appear. Instead of "Local, A-Z" the caption is "Special bases. By name, A-Z". Wording varies slightly but the pattern is the same.

CLASS T			
TL	MOTOR VEHICLES.	AERONAUTICS.	ASTRONAUTICS
TL787-TL4050		Astronautics. Space travel	
TL4000-TL4050		Ground support systems, operations and equipment	
TL4020-TL4028		Launching bases. Launch complexes Including launch facilities	
TL4020		General works	
TL4024		Planning. Location	
TL4026-TL4028		By region or country	
TL4026-TL4027		United States	
TL4026		General works	
TL4027.A-TL4027.Z		By region or state. Under each state: .x General works .x2 Special bases. By name, A-Z e.g. TL4027.F52J6, John F Kennedy Space Center, Cape Kennedy, Florida	
TL4028.A-TL4028.Z		Other regions or countries. Each country subarranged like TL4027	

Figure 10

A second type of subarrangement contained in the LCC system is the Cutter list. Typical of this are the tables of subdivisions giving "Lists of Regions and Countries in One Alphabet" and "States of the United States" which are external tables in Classes T, U and V. These lists consist of geographic names with accompanying Cutter numbers. For example, "Table of

Subdivisions, Table II, List of Regions and Countries in One Alphabet", in

Class T, includes the following:

ArgentinaA7
ArmeniaA75
AsiaA78
Asia, CentralA783
Asia, East, <u>see</u>E18
Asia, SoutheasternA785
Asia, Southwestern <u>see</u>N33
AustraliaA8
AustriaA9

A third approach to subarrangement is through notes which are technically "divided like" instructions and which are exemplified in Figure 10 in the note "Each country subarranged like TL4027." Finally, many subarrangements are provided for without tables through the use of "A-Z" instructions.

7.3.3 *Tables*

The most complex tables are the large tables which are both internal and external to the schedules. In most cases these tables are numbered (e.g. Table I, II, etc.) but not always, as can be seen in Figure 9. Tables which are not numbered tend to be located on the same pages as the schedules to which they apply, or in very close proximity. Numbered tables apply generally to whole classes or one or more subclasses and are usually located at the end of the schedule or at the end of a particular subclass. The style of numbering tends to be Roman numerals, but this is not consistent. For example in the schedule for B-BJ there are 6 large tables numbered Table 1, Table 2, etc. Table numbers apply only to a particular schedule, and sometimes only to a particular subclass. Thus it is possible to find two or more Table I's, Table II's etc. in the same physical volume of schedules. Of particular importance is the fact that large tables with spans of numbers dealing with a common topic frequently combine several tables using a common set of captions, as can be seen in Figure 9.

The means by which printed schedules are linked with the appropriate tables also varies. Small and medium sized tables that apply exclusively to one particular class number or span of class numbers are usually located adjacent to the portion of the schedules to which they apply. The link with the schedules is maintained only by the position on the printed page. Larger tables with wider application are usually linked to the schedules by a superscript (e.g. 1/) or a variety of parenthesized additions to captions (e.g. (I), (3), (Table I), etc.). The parenthesized terms identify the table number directly, while the superscripts lead to footnotes which locate the tables either by table number or page number, or both. The actual link between the schedule caption and the table varies. In some cases the precise topic to which the table is to be applied is linked to its table, while in other cases the link is made only between the generic superordinate caption and the table. However, in reality, the table is applicable to each of the subtopics subordinate to that generic caption. In this latter case, while the printed page makes this link between caption and table clear, such a link needs to be made explicit in a machine-readable version of LCC. Figure 11, from the B-BJ schedules for "Individual philosophers" exemplifies the problem.

B-BJ PHILOSOPHY

B	PHILOSOPHY (GENERAL)	B
B790-B5739	Modern (1450/1660-)	
B850-B5739	By region or country	
B2521-B3396.Z76	Germany. Austria (German)	
B2535-B3396.Z76	By period	
B2615-B2729	18th century	
B2631-B2729	Individual philosophers 1/ - Continued	
B2635	Baumeister - Baumgarten	
B2637	Baumgarten, Alexander Gottlieb	
B2639	Baumgarten - Bilfinger	
B2640-B2643	Bilfinger, George Bernard (3)	
B2644	Bilfinger - Buddeus	
B2645-B2648	Buddeus, Johann Franz (3)	
B2650	Buddeus - Creuz	
B2651	Creuz, Friedrich Casimir Carl, Freiherr von	
	...	
B2694-B2698	Nicolai, Christoph Friedrich	
B2699	Nicolai - Rüdiger	
B2699.04	Oetinger, Friedrich Christoph	
B2699.P3	Platner, Ernst	
B2699.P5	Ploucquet, Gottfried	

etc.

1/ For subarrangement under philosophers, see Tables, pp. 132-136.

Figure 11. Links Between Schedules and Tables

Disregarding the problem of node labels which is discussed on pages 25 and 26 of this report, Figure 11 above shows that "Individual philosophers" is footnoted to a general note leading to 5 tables on pages 132-136 of the schedule. This note applies to some of the names listed under the caption "individual philosophers". Exceptions, for example Bilfinger, George Bernard (3) are linked directly to their tables. Other names, for example "Creuz, Friedrich Casimir Carl, Frieherr von," for which Table 4 is applicable, as well as Oetinger, Platner and Ploucquet for which Table 5 would be applicable are only linked to their tables indirectly through the generic caption "individual philosophers". Similarly, Nicolai, Christoph Friedrich should be linked to Table 3 but the (3) is missing from the name. It may be argued that the names for which there are single class numbers or Cutter numbers are only minor

philosophers and therefore further subdivision is not needed. However, the facility is there for those libraries which may need to use it now and for the future. It is logical that every caption to which a table applies should be linked directly to its table. There are numerous instances throughout the system where the link between specific captions and their related tables are implicit in the display but need to be made explicit in the machine-readable version of the schedules.

As presently provided for in the USMARC Classification Format, in field 153 subfield code "z", tables will be linked with their particular schedules and all tables will be numbered in Roman. Does this mean that all tables will be numbered including subarrangements? Linking tables and subarrangements with their individual schedules appears to be a reasonable short term solution to the problem. However, for the machine-readable version of LCC two things should be considered, first, the consolidation of duplicate tables and subarrangements and secondly their absorption into the schedules. As long as printed schedules continue to be produced, classifiers using the printed schedules may be best served by maintaining the physical proximity of schedule data and related tables. Although here again some consolidation could be carried out.

In general, table data present the same kinds of problems as posed by schedule data with respect to hierarchical relationships, notes, references, "divided like" instructions, etc. However, the size and complexity of the large tables raise questions particular to their display and manipulation in a computerized environment.

7.3.3.1 *Screen Display of Large Tables*

Basic requirements for the use of tables in an online environment are ease of access to a table, clarity of the display, provision for viewing a table intact, provision for scanning tables which are too large for a single display and

provision for viewing table and schedule data simultaneously. It should not be difficult to satisfy these requirements where small tables and subarrangements are concerned. With the complex tables there are several questions which need to be addressed. Should the combined tables, as exemplified in Figure 9, continue to be displayed in this way? What are the requirements and the limitations if this is done? If a decision were made to handle each table separately, what are the implications of this? What are the possibilities for clear unambiguous simultaneous display of table and schedule data?

The nature of table data is such that tables which are combined in display are interdependent with respect to captions. Frequently all captions in a list common to 3 or 4 tables may not all apply to all of them, but the context and definition of the subtopics applicable to a 10 number table are useful to the classifier who is applying a 5 number table to the same subject area. Figure 9 illustrates a display of 4 combined tables. Similarly the P-PZ Language and Literature Tables have page displays of up to 5 tables across the page. It is probable that it will be neither possible or desirable to display these combined tables simultaneously on a computer screen. This will be even more so with displays of 10 tables such as those found in the "Tables of Geographic Subdivisions" contained at the end of HM-HX and Class N. A logical solution to the problem is to display only one table at a time. If this is done, the captions for the combined tables should remain intact for each table displayed separately.

It is also important to provide for the scanning of long tables by classifiers as they would if using the printed schedules. Finally, the classifier needs to be able to move from schedule to table and back to schedule with ease in applying tables in the classification process. In the light of this requirement, technology which would provide for windows or split screens and a notepad should be investigated.

7.3.3.2 Manipulation of Tables

Among their other characteristics, LCC tables and subarrangements can be characterized according to the processes by which they are manipulated in the course of manipulating the schedules. In each case a series of steps is followed in selecting the required classification number. In analyzing the tables and subarrangements, the following patterns were found:

a) The addition of a Cutter number:

Two types of procedures fall into this category. In the simplest procedure, a Cutter number is selected from an existing list, for example the "Lists of Regions and Countries in One Alphabet" and "Tables of States", external tables located in schedules such as HM-HX, U and V. The chosen Cutter number is then attached to the end of a classification number. The second procedure in this category is the application of the small tables (subarrangements) which provide for successive Cutting. Following instructions for the topic a Cutter number is selected, the successive Cutter table is examined, the correct subtopic from the table is identified and the Cutter number adjusted to match the pattern in the surrogate Cutter (.x, .x2, .x3, etc.) and attached to the classification number. In each case, further Cutting may be necessary using the LC formula for book numbers. Ultimately the result is verified against the library's shelflist.

b) Selection of a class number from a span of numbers based on its position in a sequence:

Figure 12 illustrates another frequent pattern of procedures for using tables. In this case a topic is identified by a span of numbers and a Table (e.g. Table III in Figure 12), is examined to determine the sequence of numbers in the original span which represents the topic. From there a second table (1 of 3 in Figure 12) is examined to determine which number in the sequence is

appropriate to the topic and that number is added to the number identified in the instructions. After any necessary additional Cutting, the resulting classification number is verified and adjusted against the shelflist. Typical of this process are the many cases throughout the schedules where a topical subject is subarranged "By country or region" using a table to determine the span of numbers assigned to a particular geographic entity. A variation on this can be found in the Class K schedules. The total operation increases in complexity as additional tables are added to the process, but the individual procedures do not themselves increase in complexity.

CLASS N				
NC	DRAWING.	DESIGN.	ILLUSTRATION	NC
NC50-NC257	History of drawing			
NC101-NC376	Special countries. Table III 1/ Prefer classification by subject except special artists			
	Under each (three number countries):			
	(1) General works			
	(2) Local, A-Z			
	(3) Special artists, A-Z			
	Under each (one number or decimal number countries):			
	.A1A-Z General works			
	.A2-Z8 Local, A-Z			
	.Z9A-Z Special artists			
	Under each (Cutter number countries)			
	.x General works			
	.x2 Local, A-Z			
	.x3 Special artists, A-Z			
	...			
<u>1/</u> For Table III, see pp. 345-353. Add country number in table to 100.				

Figure 12. Tables and Subarrangements

c) Selection of a classification number based on tables using terminal digits.

A third, frequently used pattern, found in some of the more specialized schedules, is one where the basis for selection is one or two terminal digits in

each class number in a span. Typical of this type are "Tables of Subdivisions" for "Colonies and Colonization" in Subclass JV, Tables 1 and 2 used in subarranging topics under "Individual philosophers" in BQ, and several tables in Class P. A span of numbers representing the main topic is compared with the appropriate table and on the basis of a book's content a subtopic is selected and the table numbers are matched against the terminal digit(s) in the number span to select the required class number. In some cases it is necessary to make a choice of sequences as shown in the small table below.

Subarranged:

Works with 5 numbers:

0 or 5	Original texts
	By date
1 or 6	Editions with
	Commentary
	By editor
2 or 7	Selections.
	Paraphrases, etc.
	By date

While the sample may not have identified all the types of tables and subarrangements present in the LCC system, the significance of the three types identified above is that a number of definite patterns of procedures exist which require "identification", "selection", "comparison" or "matching", "adding to" "attaching to", etc. Individually each procedure could be described in a simple command either to the computer or to the classifier. Moreover, there is significant repetition of these procedures in groups of tables across the schedules. These are characteristics which lend themselves to the development of an expert system, or group of such systems, using a command-driven mode to permit the classifier to interface with the system making the necessary intellectual decisions while the computer carries out the mechanical processes. To be most effective, such a system needs to include automatic application of the formulae for "Book Numbers" and automatic comparison with an online shelflist. If such a system were developed, it would remove some of the mystery and much of the tedium associated with manipulating the tables when using the printed schedules.

7.3.3.3 Summary of Findings

Clearly, the tables will be one of the most difficult components of the LCC system to deal with in terms of display and manipulation. Could tables be abolished? It might solve many problems, but create others. Based on the evidence from the analysis, the answer to this question seems to be "Yes, but probably not immediately". There is strong evidence that it could be done with time, much patience and the resources to do it. Some of the tables are open-ended in the topics they cover. That is, all of the topics covered by the tables are not predictable. However, this in itself is not an obstacle because new topics can be added to the schedules at any time. A major problem would be the amount of conversion that would be required and the number of individual intellectual decisions that would have to be made in the process of amalgamating the content of the tables into the main schedules.

Time did not permit a complete and accurate assessment of the total impact of amalgamating schedule and table data, but a brief analysis to two tables, one from each of Class N Fine Arts and Subclass KE Law of Canada, gives some indication of the volume of work which would need to be done. Using the KE schedule, Table I was examined for its impact on the schedules. This table is identified as being applicable to spans of 10 numbers. In reality, the Table includes 10 numbers plus subarrangements under those numbers identified by decimals and/or Cutter numbers, totaling 70 notations. When the node labels and see references, each of which represents a USMARC record are added, the number 70 changes to 84. There are 30 occurrences of the use of Table I, or Table I "modified", in KE, which would add 2,520 records to the KE schedules, and of course there are several other tables which would need to be integrated also. It should also be noted that this is a table of "Form Subdivisions" including such subtopics and "bibliography", "serials", "citators" etc. It is a very straightforward table without any special arrangements and the numbering system is based on terminal digits in the class numbers. Integration of table and schedule records could be quite easy in this case.

Following the same procedure for analysis, Table II in Class N was analyzed for impact on the schedules. Table II in this schedule is a table of geographic subdivisions and is applicable to continents and countries which have been assigned 200 numbers. It contains captions, see references and examples with Cutter numbers attached. Also there are 2 footnotes, one of which is a subarrangement to be used with individual states of the United States, and the other provides a subarrangement for Scotland, Ireland and Wales. The number of captions and see references in the main table represent 238 records. Added to this would be the 59 records from the two subarrangements for a total of 297 records. There were 8 occurrences of the use of Table II across the N schedule representing 2,376 records to be integrated. In addition all of the occurrences of Table II lead to a choice from 3 "second level" subarrangements at N101-N284, for 2 number countries, one number or decimal number countries, or Cutter number countries.

From these two examples it is not possible to do more than speculate that abolishing most or all of the tables might be possible. A move in this direction should be seriously considered in the future. Almost anything is possible in the rapidly changing technological climate of the future. However, if the need is to convert the schedules as quickly as possible, now may not be the time to consider this possibility. A more thorough study is needed before such a step is taken. For the immediate future, there are some adjustments which are essential and some which could be made which would make the tables and subarrangements easier to use and maintain.

It will be essential to ensure that tables and subarrangements are directly linked to the captions to which they apply, a uniform numbering system provided, and footnote instructions integrated into the schedule data. Generic links with superordinate captions and footnotes will not be viable in the computer environment. The screen display is also crucial. It should permit classifiers to scan the tables and to have simultaneous viewing of the schedules

and tables together with a notepad for building class numbers. The consideration of expert system manipulation of the tables should also have a high priority.

Since some duplication exists it would also be possible to consolidate some tables, particularly the external geographic tables. At the very least it should be possible to use one set of Cutter lists for geographic names across schedules. There is much duplication in the small internal tables which provide for successive Cuttering. However, they are not all exactly alike and would require careful scrutiny. If they were to be consolidated for the computerized schedules an important consideration would be how to handle these small tables in the printed schedules. Physical location in a computerized system is irrelevant. However for convenience of use, in the printed volumes there is a certain virtue in having access to these subarrangements on the same or adjacent pages as the topics to which they relate.

7.3.4 RECOMMENDATIONS

In the light of these findings it is recommended that:

- a) Where possible tables and subarrangements be consolidated and reduced in number; the geographic Cutter lists are obvious candidates as "possibly" are some successive Cutter subarrangements;
- b) A consistent style of numbering of tables be adopted;
- c) The integrity of the relationship among tables which have a common set of captions be maintained (e.g. figure 9);
- d) The links between captions and tables be edited so that each caption which requires the use of a table is linked directly to the appropriate table, rather than indirectly through a generic caption superordinate to it;

- e) Further study be made of the display requirements needed for large and complex tables giving consideration to the need for simultaneous display of schedule and table data and the ability to scan tables too large to display on a single screen;
- f) The possibility of expert system(s) (for manipulating tables and subarrangements) be investigated so that the intellectual decisions are made by classifiers, while the mechanical processes are carried out by computers;
- g) Any differences in the display requirements for printed and machine-readable versions of LCC which will affect the USMARC format be identified; and
- h) Thought be given to further research which sometime in the future could lead to the elimination of some or all of the tables.

7.4 Notes and References

7.4.1 General Characteristics

The LCC schedules contain many notes and references which serve to explain or define topics, to provide instructions for classifiers on how to manipulate the schedules and tables, and to link topics and aspects of topics located in widely separated parts of the schedules. Most notes and references are integrated into the schedules and tables. However, some appear as footnotes, while still others apply generally to a class or subclass and are located at the beginning of the class or subclass to which they pertain.

Any particular topic may have one or more notes attached. When there are several notes and references, the format varies. Each note may be presented as a separate paragraph, or they may be grouped together in a single paragraph.

The notes in Subclass HM-HX at "HT101-HT395 Urban groups. The city. Urban sociology", as shown below, demonstrate both formats.

Class here works dealing generally with the city as a social group. For special aspects of the city, see HT or elsewhere as indicated below

Cf. GT3420+, Town life (Manners and customs)

For individual cities, see D-DU, HC, HN, J, JS, etc.

For urban geography, see GF125

For urban anthropology, see GN505

The first two notes form a single paragraph, while the Cf. note and the 3 "For ... see ..." references each form a separate paragraph. In some schedules, a single note may be given and bracketed to 2, 3 or more captions indicating that the same note applies to each of the captions in the bracket. In other cases, notes were found to be extensions of the captions. For example in the caption "Early (To 1800). By date", "Early (To 1800)" is the caption. "By date" is really a note requiring subarrangement. Similarly in "Unauthorized negotiations (including works on their criminal aspects)", the addition to the caption in parenthesis is actually an "Including ..." note while "A-Z" is a subarrange instruction. However, in LCC there are exceptions to almost every rule. In captions such as "Fuses (General)", "Tile (Ceramic)" and "Moldings and cornices (Interior)" the parenthesized terms are indeed part of the captions. For purposes of conversion to machine-readable form, and in consultation with LC staff, it was considered essential to categorize the data according to its function. Therefore in this study, notes which were formatted as extensions to captions were categorized as notes and each note was counted as a separate note regardless of the format.

As could be expected, notes and references vary in nature and format across the system. Some notes are unique; some are unique to a particular class or subclass. Nevertheless a significant number of notes and references have a discernible pattern in wording, format and function which has been applied across all schedules in the LCC system. In the process of this study

12,504 "see" and "for ... see" references were located in the sample of 3,934 pages, together with 11,462 notes and 656 footnotes, for a total of 24,622 notes and references, that is 6.3 of these devices per page of schedules and tables. From this, nine distinctive types of notes and references were identified as follows:

- a) "... see ..." references
- b) "Cf. ..." notes
- c) "Including ..." notes
- d) "Class here ..." notes
- e) "For ... see ..." notes
- f) "Prefer ..." notes
- g) "Divided/Subarranged like ..." instructions
- h) "Subarranged by ..." instructions
- i) "Subarrangement of captions (i.e. "A-Z" instructions)

In addition a wide variety of "other" types of notes were identified from the sample. Some of these also lend themselves to categorization, although separate counts were not recorded. For example there were notes which included the words "class ... in ..." and "class ... with", both of which might be described as "class elsewhere" notes. As well there were numerous examples preceded by the abbreviation "e.g.". These examples varied in nature and function. Some functioned as scope notes, or partial scope notes, as for example "e.g. Freemasonry and international law" located at "HS499 Other special" under "Societies: Secret, Benevolent, etc. Freemasons". These were categorized as notes. Others, such as "e.g. .C6 Commonwealth Fund", were attached to Cutter numbers. Since this is the same type of data as found in the many lists of Cutter numbers throughout the schedules, examples attached to Cutter numbers were categorized as captions and included in the hierarchies. Some other e.g. notes were "see" references and were categorized as such. Typical of the notes which were unique to a particular class were the history notes in LG and the history and "biography" notes in E-F. Other notes and references were really instructions on how to use the schedules and tables, as for example the instruction at HV742 which directs the classifier "For tables of regions, see p. 208".

In addition to the general characteristics of notes and references, each individual type of note and reference has its own characteristics and raises its own problems.

7.4.2 See References

Three types of references that describe themselves as "see" references were found in LCC. These include "see" references which lead from locations in the schedules where no classification numbers are provided for topics to classification numbers for those topics. They are "independent" in that in the location referred from they are not physically attached to classification numbers. Other "see" references lead from classification numbers formerly used for topics to new classification numbers which relocate the topics to other places in the schedules. The third type of "see" reference ("for...see...") leads from a valid classification number and its caption to aspects of that topic or closely related topics in other parts of the schedules. Each of the three types of "see" references has its own particular characteristics and functions.

The primary purpose of all "see" references is to direct users of the schedules from a topic at one particular location in the schedules to one or more other locations of that topic, aspects of that topic and/or related topics. In doing so, these references compensate for the multidimensionality of knowledge and the linearity of the classification schedules. It is essential that relationships be established among topics and subtopics which are otherwise scattered by the classification system. To support this purpose effectively, the references must be easy to use and provide users with a clear understanding of the topical environment in which each reference is operating. Therefore it is essential not only that the references exist, but also that they function properly. While this is important when using the printed schedules, it becomes even more crucial in the computerized environment where the relationships across the system becomes more difficult to visualize.

In this study, in addition to collecting statistical data, it was necessary to assess the possible effectiveness of each of the three types of see references in an online environment and to identify problems which could arise for classifiers and other users navigating LCC online. A detailed examination was carried out on a random sample of each type of reference taken from each LCC schedule. Each of the three types of references was analyzed, using references which were internal to a single schedule and references which referred to class numbers in other schedules. Each reference was evaluated for:

- a) Its viability (i.e. to ensure that it was not a blind reference);
- b) The degree of similarity between the terminology in the reference and the terminology in the topic referred to; and
- c) The orientation to the topic referred from with its new location.

The sample of 3934 pages from the LCC schedules revealed a total of 15,787 "see" references of some kind, including 9,654 "see" references and 3,304 "for ... see ..." references, including footnotes. The average number of references per page was 2.45 "see" and .83 "for ... see" references. Based on these data it is estimated that there are 31,077 "see" references and 10,636 "for ... see" references in the LCC system. The density of the "see" references varies considerably across the schedules. The range is from 1 to 4 references per page, depending on the subject matter and, probably, on the perspectives of the editors of the various schedules. The highest density was found to be in Classes K and P. In Class K, Law, the schedules for subclasses KDZ and KG-KH averaged 4 "see" references per page, while subclasses KJV and KK had densities greater than 3.5 references per page. The only other schedule with a density of 4 references per page was PT, Part 1 German Literature, but schedules E-F, PG, PJ-PM and PQ, Part 2 also revealed averages greater than 3.5 "see" references per page. At the lowest end of the density scale were the M and U schedules and the P-PZ tables, all of which contained less than one "see" reference per page. The remaining schedules fell between 1 and 3.5 references per page.

Since the "for ... see ..." references represent only approximately 21% of the total number of references in the sample, their density is much less significant. Subclasses DJK and KE had an average of 2 "for ... see ..." references per page; 20 schedules had less than 1 and the remainder less than half a "for ... see ..." reference per page.

SUBCLASSES HM-HX				
HT	COMMUNITIES.	CLASSES.	RACES	HT
HT101-HT395	Urban groups.	The city.	Urban sociology	
HT165.5-HT169.9	City planning			
HT169.6-HT169.9	Zoning			
HT169.9.A-HT169.9.Z	Special topics.			
HT169.9.D4	Development rights transfer			
HT169.9.E82	Exclusionary zoning			
	Housing, <u>see</u> HD7285+; NA7511+			
	Slums, <u>see</u> HV4023+			
	Settlements, <u>see</u> HV4175+			
	Public works, <u>see</u> HD4421+; TD21+			
	Public buildings, <u>see</u> JS201+; NA4201+; 4430+			
	Streets, <u>see</u> HE369+			
	Parks, <u>see</u> SB481+			

Figure 13. See References

7.4.2.1 Independent See References

The first of the three types of "see" references can be described as "independent" or "true" see references. As illustrated in Figure 13, "independent" see references refer classifiers and other users from a location in a particular hierarchy in a schedule, where they might expect to find a particular topic to a preferred location elsewhere in the schedules. See references of this type can be found in hierarchies of topics in the main schedules, in the tables and subarrangements and in the lists of Cutter numbers which are frequently part of the main LCC Schedules. As can be seen above, these references are "independent" only in the sense that they are not attached to classification numbers. Nor are they part of the record for the preceding caption. They are, however, context-dependent on the hierarchies in which

they appear. For example, in Figure 13, "Housing" and "Public works" must be understood in the context of "Communities. Classes. Races" and "Urban groups. The city. Urban sociology." At the same time, it can also be seen that see references can be hierarchically related to each other and be both superordinate and subordinate to other references. Since there are no classification numbers to permanently "fix" these references in their prescribed locations, their positions in the schedules are defined only by their relationships with other references and topics that surround them on a printed page. Moreover, they are only meaningful when they can be viewed in their proper context. That it is essential that this context be preserved and made explicit in the computer environment has already been recognized in the provisional USMARC Format for Classification Data through the use of "dummy" or "invalid" classification numbers in the 453 field.

The format of these references is relatively consistent. Each see reference is composed of a caption and the word "see" followed by one or more classification numbers. Individual classification numbers are frequently followed by a plus (+) sign, signifying that a span of numbers is applicable to the topic. Some references are very generic in nature and lead to a whole class or subclass. No captions are given for the classification number(s) referred to, even though those captions may differ in terminology and level of specificity from the caption referred from. Typical of the format are the following examples:

Civic reform, see JS
Municipal corruption, see HV6310; JS231
Municipal government, see JS
Public buildings, see JS201+; NA4201+; 4430+
Women in industry, see HD6050+
Women in social reform, see HN49.W6

As can be seen from the reference "Public buildings, see JS201+; NA4201+; 4430+", some of the numbers in the references are not complete. The "NA" in "4430+" is understood, rather than being explicit. Similarly, Cutter lists sometimes contain see references which are internal to the list as

shown in the partial list taken from "BJ1533 Special virtues, A-Z" in Figure 14. Here only the Cutter numbers are made explicit, while in reality they are attached to the classification number for the caption which is superordinate to the list. The complete numbers are implicit on the printed page, but must be made explicit for computer searching.

SUBCLASSES B-BJ		
BJ	ETHICS	BJ
BJ1518-B1697	Individual ethics. Character. Virtue ...	
BJ1533.A-BJ1533.Z	Special virtues. ...	
BJ1533.D49	Discipline. Self-control ...	
BJ1533.G6	Glory Good workmanship, <u>see</u> .W6	
BJ1533.G8	Gratitude ...	
BJ1533.R42	Respect for persons Self-assurance, <u>see</u> .S27 Self-confidence, <u>see</u> .S27 Self-control, <u>see</u> .D49 Self-reliance, self-assurance, self-confidence ...	
BJ1533.S27		
BJ1533.W6	Workmanship, Good	

Figure 14. Cutter List With References

7.4.2.2 See References Which Relocate Topics

The second type of "see" reference, references which relocate topics in LCC, are a result of changes made in the classification system over the course of its history. These references refer from classification numbers which are no longer used to classification numbers currently used for the same topics. The older classification numbers have been retained because LC, and possibly other libraries, continue to have parts of their collections located in these numbers. In effect, they are "alternative" locations for some topics in the system and, technically, these references function as "see also" references. For purposes of this study, statistically they were handled as "independent" see references. References of this kind can be found unevenly distributed throughout the

schedules. Some schedules have none; others have many. For the most part they are located in schedules with subjects where major changes in policy have taken place.

One of the best examples of the use of "alternative" classification numbers can be seen in conjunction with topics in "Law", which were previously distributed across disciplines but have been relocated in the recently developed Class K schedules. For example, at the beginning of Class J, Political Science, there is a general reference "Gazettes, see Class K".

As exemplified below, the format for this type of "see" reference includes a classification number in parenthesis (to indicate that it is not currently being assigned to documents), a caption, the word "see", and the classification number currently used.

JC(117)	The City State, <u>see</u> JS61, etc.
JK768(.5)	Bonding of government employees, <u>see</u> HG9970.S5
JK(645)	Relation of the departments to Congress, <u>see</u> JK585
JK717(.S7)	Stenography, <u>see</u> Z49.2; 53-102

These references can be found in both schedules and tables where they apply to either a single class number, or spans of numbers. Where there are many consecutive "alternative" numbers, there may be a single generic reference at the beginning of a class or a section of a class rather than references attached to each individual topic affected, as can be seen in the example from Class J in Figure 15. The reference "Gazettes, see Class K" functions both as a generic see reference to class K and as a node label for all the captions subordinate to it in Class J. This particular reference applies to seven and one-half pages of geographic name captions, but the reference is not linked directly with any of the individual captions.

CLASS J		
J	GENERAL LEGISLATIVE AND EXECUTIVE PAPERS	J
J(1.A1)-J(1.A4)	Gazettes, <u>see</u> Class K	
J(1.A5)-J(1.W)	United States	
J(2.A1)-J(2.A4)	States	
J(2.A5)-J(2.Z)	Canada	
	Provinces (including Newfoundland)	
	...	

Figure 15. General References

Other "alternative" classification numbers identify special subarrangements which are provided for internal use at LC. In such cases, there are no alternative numbers in the schedules. Therefore there are no references. An example of such a special arrangement can be found at "HV(2250)-HV(2329) Library of Congress collection of books for the blind".

As long as these parallel locations continue to be used for some topics, these "relocation references" must remain in the schedules. They will not cause any particular problems for an online system. However, any generic references will need to be linked to each of the captions to which they apply.

7.4.2.3 "For ... see ..." References

The third type of "see" reference found in the schedules is the "For ... see ..." reference. Technically, these references function as "see also" references. They may appear in general statements at the beginning of classes or subclasses, under node labels, or attached individual captions. They may be generic in nature, or they may refer to specific classification numbers. The format used is fairly consistent. Each reference begins with the word "For" followed by a statement of the topic referred from, the word "see" and the classification number referred to. The caption for the number referred to is not included. Exceptions to this are "For ... see" references that are generic in nature such as "For works on subjects taught, see the subject, e.g. HQ763,

Family planning; HQ768+, Child rearing". "For ... see" references are of two kinds, those which handle classification data and those which provide instructions for manipulating the schedules. "For ... see" references which contain classification data refer from a valid topic with a valid classification number to aspects of that topic which are located elsewhere in the schedules. As such they function as "class elsewhere" notes. In this capacity they often accompany "class here ..." notes. In such cases sometimes examples are included to help users to focus on the kinds of "subjects" they should consider. However examples are not included in all such cases. Typical are the following examples found at "HV1450-HV1493 Aged":

Aged

...

For blind aged, see HV1597.5

For contributory old age pensions, see HD7105+

For developmentally disabled aged, see HV3009.[sic]6.A35

For private medical care plans, see RA413.7.A4

For private old age insurance, see HG9426+

For public medical care plans, see HD7101+

In contrast, "For ... see" references, which are instructions, carry no classification data. At HV742, for example there are two such notes which refer the classifier to tables: "For tables of regions, see p. 208" and "For tables of states, see pp. 209-210". Nearly all notes of this type carry page numbers, which like other occurrences of page numbers will need to be replaced by something more precise for the online system.

7.4.2.4 Problems With References

When the performance of each of these three types of references was examined more closely it was found that references leading to other locations in the same schedule were relatively reliable insofar as they almost always led to the classification number which was identified in the reference. In contrast, approximately 10% of the sample of references leading to classification numbers in other schedules indicated problems. Some references were blind

references; many led to numbers which were "hidden" in spans of numbers and required the use of tables to locate the actual number. In other cases there were errors. For example at "HV1450-HV1493 Aged", "For developmentally disabled aged, see HV3009.5.A35" has a typographical error in it. The topic referred from is at HV3009.6.A35. There is no class number HV3009.5 in the Gale 1986 edition. In another example, one of the classification numbers referred to was technically available but was not used in the schedules. With the "see" reference "AE(90) Other, see AE5-89.9", AE89.9 did not exist in the schedule, although it could be used for a topic in the future. AE88 was the last currently valid number in the span. In spite of this the number of such errors was remarkably few considering the volume of references in the schedules and the lack of tracings for these references under headings referred to - a device which would aid the editing. Indeed the number of errors was smaller than was predicted from the pre-test. In part, this may be due to the fact that most of the schedules finally used in the project came from a uniform edition of the schedules which was complete with "additions and changes through" the same date (Gale 1986). Many of the schedules used in the pre-test were published at different times and would not have been updated with all the changes which had taken place. Nevertheless the provision for tracings in fields 453 and 553 in the USMARC format will do much to ensure that consistency is maintained.

Effective use of the schedules is dependent not only on the presence of references to navigate the schedules, but also on the effective performance of those references. Among other things this means ease of use and a clear understanding of the references in the context of the schedules. For purposes of evaluating their performance, references of the three types were randomly selected from the pages of the original sample, and each reference was followed through to the classification number to which it referred. A number of problems were encountered which could hinder use of the schedules and, at the very least, cause frustration for the online user.

Since many of the references refer to classification numbers without

including the captions of the classification numbers referred to, the user is dependent upon being able to understand the topic referred from in its new context in some other location in the schedules. This is not always easy. When the exact class number is located, the precise caption may contain the exact terms which were in the first part of the reference or the caption may be quite meaningless on its own, as for example "General works", or "Periodicals". The hierarchical context is there to help with orientation to the new context. However, if the hierarchy is complex or the reference involves "upward posting" of a very precise topic to be included in a more general topic, the term which the user started with in the reference may be nowhere in sight in the new location. For example the reference "Municipal corruption, see HV6310" (following HT281) leads to "HV6310 City and state 'rings'". By definition the topics are related, but how easy is it for the user to make a quick mental connection? As shown below the topic is 9 levels down in its hierarchy:

SOCIAL PATHOLOGY. SOCIAL AND PUBLIC WELFARE. CRIMINOLOGY
Criminology
 Crimes and offenses
 Political crimes
 Offenses against the administration, election laws, etc.
 By region or country
 United States
 Administration
 Bribery of officials. Veniality in office
 Graft
 City and state "rings"
 Land steals

Nearest neighbours held to define the topic, but it is not as effective as being told directly that "Municipal corruption" is included here.

SUBCLASSES HV-EK

HV SOCIAL PATHOLOGY. SOCIAL AND PUBLIC WELFARE. CRIMINOLOGY HV	
HV697-HV4959	Protection, assistance and relief
HV697-HV1493	Special classes
HV701-HV1420.5	Children
HV851-HV861	Day care centers. Foster day care
HV854-HV861	By region or country
	...
HV862-HV866	Residential care. Group homes
	...
HV867	Baby farming
HV868	Milk depots
	Cf. HD4501.M5 Municipal industry
	LB3473+ Feeding of school children
	Children's playground, <u>see</u> GV421+
	Chimney-sweeps, <u>see</u> HV873+
	Blind, deaf and dumb, <u>see</u> HV1551+
HV873-HV887	Destitute, neglected, and abandoned children.
	Aid and homes
HV873	General works
HV874.8-HV875.7	Adoption
HV874.8-HV874.82	Biography
HV874.8	Collective
HV874.82.A-HV874.82.Z	Individual

Figure 16. See References

Figure 16 contains 3 see references which further illustrate the problem.

The reference "Blind, deaf and dumb, see HV1551+" leads to "HV1551 Periodicals. Societies. Serials.", which is subordinate to the node label [HV1551-HV3024] "Handicapped". "Handicapped" in turn includes a note "Including general works on the blind, deaf, mentally handicapped and physically handicapped." Using the proposed USMARC record display, the record for HV1551 will show the topic "Periodicals. Societies. Serials" in the context of the hierarchy which includes "Handicapped". However, that record will not show the useful "including ..." note. That note will be in the record for "[HV1551-HV3024] Handicapped". Also the classifier will not know from either of these records that "Blind" is included in [HV1571-HV2349] and "Deaf", including deaf-mutes and hearing impaired, will be located in [HV2350-HV2990.5] In another example, "Chimney-sweeps, see HV873+ leads to "HV873 General works", which has as its superordinate topic

"Destitute, neglected, and abandoned children. Aid and homes." The term "Chimney sweeps" appears nowhere in the hierarchy. The tracing of the reference in the USMARC record will help here. On the printed page "Chimney-sweeps, see HV873+" is two lines above "Destitute, neglected, and abandoned children...". However there is a lack of clarity which could be confusing in the online display. "Children's playgrounds, see GV421+" presents a similar problem. It leads to "GV421 Periodicals. Societies. Serials" whose superordinate heading is "Playgrounds and play spaces".

What is happening here? In cross-checking more than 400 references, it was discovered that see references lead to classification numbers which are visible on the printed page and not to captions. Where a reference leads to a single class number representing a precise topic and that topic has no subordinate topics of its own, the caption at the location referred to will usually be a descriptive caption. For example, following TP897 in the schedules, the reference "Dyeing of cotton, see TP930" leads to "TP930 Textile printing. Calico printing. Cotton dyeing". However, there are at least three types of situations where references lead to non-descriptive captions. In the first type, references have been made directly from one non-descriptive caption to another, for example "Addresses, essays, lectures, see Minor works". Such references are entirely context dependent on the hierarchy referred from and the hierarchy referred to. The other two types are both related to the page display in the schedules, where the "real" descriptions are node labels representing spans of numbers which are implicit in the displays but have not been made explicit on the printed pages. In these cases the only visible evidence is a class number which is attached to a caption which is subordinate to a node label.

From the above examples it is easy to see how some confusion and frustration may arise in the course of using various kinds of see references. What are some possible approaches to solving this problem? All see references could be edited to add the captions of the class numbers referred to. Since provision is made in the USMARC record for making explicit in the records

the spans of number which apply to the node labels, the references could be edited, where appropriate, to refer to the number spans and their captions. Given the number of see references in the schedules this would require a massive editing program for the best results. Nevertheless it does not solve the problem entirely. The hierarchical context is still necessary. Regardless of the kind and degree of editing possible, two things will be important. The system must permit the user to move from the location of a reference to the place referred to and to be able to browse the relevant area in the schedules. Secondly, it would be helpful if the system were to permit simultaneous viewing of a reference in its hierarchical context and the topic referred to in its hierarchical context. Some kind of windowing device supported by a notepad appears to be needed.

7.4.3 Notes

The LCC schedules contain numerous notes. Many conform to one of several types; others are unique in content and format. Seven specific types of notes, which occur frequently throughout the schedules and which are formalized as to content and format, were located in the sample. Four of these types carry classification data, while three were instructional in nature. Most types of notes can be found in both schedules and tables. Notes carrying classification data provide links among topics across the system and define classes, subclasses, topics and subtopics. They are context dependent and must be understood in the context not only of the captions under which they appear, but also in terms of the hierarchies above them. Some notes also apply to the topics and subtopics which are subordinate to the captions to which they are attached. Two types of notes, "cf. ..." notes and "prefer ..." notes actually function as cross references in the form of "class elsewhere" and "class here" notes, while "including ..." notes serve as scope notes, or partial scope notes. The three types of instructional notes provide for the extension of the schedules through "divided like" and "subarranged" instructions and are discussed later in this report in Section 7.7 on "Divided Like Instructions". Notes unique to a

particular caption and unique to a particular schedule or subject area were found in small numbers. For example some LCC classes lend themselves to history notes.

7.4.3.1 "Cf. ..." Notes

"Cf. ..." notes occur in both schedules and tables, and in some cases apply globally to large portions of a class or a subclass. They are always attached to a caption of some kind. This may be the caption for a valid class number, a node label representing a number span, or the heading for a class or subclass. "Cf. ..." notes refer from the captions, or topics, under which they appear to aspects of that topic or related topics which appear elsewhere in the same, or different, schedules. Most cf. notes refer to a single class number, a string of individual class numbers, or a number span represented in the schedules by a node label. Others are generic in nature and lead generally to subjects or to a whole class or subclass. Some examples of cf. notes are the following:

LB3431 Strain and overpressure. Mental capacity
Cf. LB1075, Fatigue

[HV8593-HV8599] Torture
Cf. K. Law

[HV5215-HV5218] Biography
Cf. HV5301+ Alcoholism

In the first example, LB3431, "Cf. LB1075, Fatigue" is typical of many of the simplest types of cf. notes found in LCC. This note contains a single classification number together with its caption and the note leads to an exactly matching entry in another part of the schedules. Other cf. notes are somewhat more complex. In the second example above, [HV8593-HV8599], "Torture" is a node label with an implicit span of numbers and the cf. note, Cf. K, Law is generic, referring generally to the whole of class K and its subclasses. Although there may be topics related to "torture"

in all 7 of the K schedules analyzed, a search of their indexes located references to torture in only 3 of the schedules. Subclasses K Law (General) and KD Law of the United Kingdom and Ireland had index entries under "Torture (Criminal procedure)", while Subclass KK-KKC Law of Germany had an entry under "Torturing animals (Offenses under the public order)". Subclasses KE, KF, KJV-KJW and KDZ, KG-KH had no similar index entries.

Other cf. notes lead to class numbers, where the caption does not match. Such is the case with "[HV5215-5218] Biography". Here the note "Cf. HV5301+ Alcoholism" leads to "HV5301-HV5720.5 Other regions or countries. Table IX 1/", which appears in the following hierarchy:

SOCIAL PATHOLOGY. SOCIAL AND PUBLIC WELFARE. CRIMINOLOGY
Alcoholism. Intemperance. Temperance reform
By region or country
Other regions or countries

As the example indicates, the cf. note leads to the first number in a range of class numbers and not to the caption "Alcoholism" which is the node label, two levels above in the hierarchy. This reveals a discrepancy between the data in the note and the data at the location referred to in the schedules. This is a problem analogous with the problem revealed in the analysis of the references.

Cf. notes perform two basic functions. They alert classifiers and other users to other classification numbers which may be useful to consider in the act of classifying and/or the search process. Also they serve as gathering devices for bringing together aspects of topics and related topics scattered throughout the schedules. As such, they function as "see also" references and are the related term (RT) relationships of the LCC system. In this respect their function is very close to that of "for ... see ..." references. Indeed, numerous instances in the schedules suggest that the functions of these two devices, the "for ... see" references and the cf. notes, are not always differentiated in the LCC system.

There are many cf. notes. The sample used in the study contained 6233 cf. notes, an average of 1.6 cf. notes per page. Based on these figures, it is estimated that, potentially, the LCC system may contain 20267 cf. notes. Many pages, particularly lists of names of persons or institutions, have only a single cf. note attached, while others, for example "[HQ1060-HQ1064]. Aged. Gerontology (social aspects) . Retirement" have long lists of such notes. The largest concentration of cf. notes was found to be in Class R and Subclass KJV, both of which revealed an average of 4 cf. notes per page of schedules.

The format of cf. notes varies from schedule to schedule, two distinct patterns of format being used. Each single cf. note, or first note in a list is preceded by the letters "cf.", meaning "confer". In one format each note consists of a classification number, frequently followed by a plus (+) sign representing a span of numbers, and a caption. This caption may be the same caption which is attached to the classification number at its own location in the schedules, but as discussed later, frequently this is not the case. In the second format, the caption is omitted from the note and only the number is given.

A single cf. note may refer to one or more classification numbers and separate notes under the same caption are usually displayed as a list. The classification numbers are usually complete, including the class or subclass letters and, where they are part of the schedules, relevant Cutter numbers. However, where the cf. note consists of a string of numbers from the same class or subclass, the class and subclass letters are included in the first class number in the string and omitted for the rest.

As is the case with references, cf. notes are not traced under the class numbers to which they refer, making maintenance and editing time consuming and error prone in maintaining the printed schedules. Examples of the formats are:

[HB848-HB3697] Demography. Vital events

Cf. HQ750+, Eugenics

RA407+, Medical statistics

HE6499

Civil service. Appointments. Personnel. Compensation

Cf. HE6161, Mail carriers

HE8475, Railway mail service

[PD2401-PD2477] Modern Icelandic language (ca. 1550-)

PD2409

History of language

Cf. PD2225

Very few "blind" cf. notes were found in the checking the notes in the schedules, when the schedules were from the same edition of LCC (i.e. Gale, 1986). A small number of mistakes in the class numbers were found. Most of these could be attributed to typographical errors but the number of such errors was minimal in comparison with the volume of cf. notes found in the schedules. Nevertheless a subset of the cf. notes selected from the sample revealed other significant technical problems. Not unexpectedly, these problems were similar to those identified in conjunction with "see" and "for ... see" references. Specifically the cf. notes frequently lead to classification numbers rather than to the precise caption to which they refer.

As with the references, it is essential that the cf. notes be clearly understood in the context of their location in the schedules. In many cases direct relationships are clear, simple and straightforward, as in:

NC773

Sketching of portraits

Cf. HV8073.4, Police composite drawings

NC860, Crayon portraits

NK7300-NK7399 Jewelry

Cf. GT2250-GT2280, History of costume

TS740-TS761, Jewelry manufacture

TT212, Jewelry craft

These examples illustrate the way in which cf. notes serve a gathering function by collocating aspects of topics from across schedules. These particular examples illustrate cf. notes which lead to specific class numbers with precise captions. However, many cf. notes which refer to spans of numbers identified by the plus (+) sign lead to the first occurrence of the first number in the span which is frequently attached to such captions as "General works", "Histories", "Minor works", "Periodicals, etc." Some cases were also found where the caption at the number referred to was itself subordinate to another non-descriptive caption. The following example illustrates this problem:

HS	Societies: Secret, Benevolent, etc.
[HS351-HS929]	Free masons
[HS473-HS475]	Antimasonic literature
	Cf. HS181+, Anti-secret-society literature
	HS525+, Antimasonic controversy, 1827-1845

leads to

HS	Societies: Secret, Benevolent, etc.
[HS101-HS330.7]	Secret societies
[HS181-HS191]	Anti-secret-society literature
[HS181-HS183]	For antimasonic, <u>see</u> HS473+, or HS525+
HS181	Periodicals
	American and English

As indicated by the square brackets the first occurrence of "HS181" on the printed page is at the caption "American and English", a non-descriptive caption, which is itself subordinated to another non-descriptive caption "Periodicals". On the other hand, the cf. note refers to the superordinate caption "Anti-secret-society literature" for the number span "[HS181-HS191]" which is not made explicit in the schedules. On the printed page the relationship is easily understood, but it could create confusion in the online system. This same problem occurs in conjunction with cf. notes leading to single class numbers wherever there are node labels involved. The caption

"General works" is particularly troublesome in this respect. For example, in Subclass LB the caption "Reading (General) reveals the following:

LB	Theory and Practice of Education
[LB1025-LB1050.7]	Teaching (Principles and practice)
[LB1049.9-LB1050.7]	Reading (General)
	Cf. BF456.R2, Psychology of reading
	LB1525, Reading (Primary education)
	etc.

leads to

LB	Theory and Practice in Education
[LB1501-LB1547]	Primary education
[LB1523-LB1544]	Branches of study
[LB1524-LB1544]	Special
[LB1525-LB1525.8]	Reading
LB1525	General works

Here it can be seen that the cf. note "LB1525, Reading (Primary education)" leads to "LB1525 General works" which is subordinate to "Reading". This reveals two problems. The caption referred from does not match the caption at the number referred to or the caption which is immediately superordinate to "General works". Moreover, the term "Primary education" only appears in a superordinate position 4 hierarchical levels above "General works" and at 3 hierarchical levels above "Reading". It is apparent from this that in terms of conversion to machine-readable form virtually every cf. note needs to be checked to make certain that both its content and context are understandable to users.

A similar situation can be found where both "for ... see" references and cf. notes are involved. For example at LB1045 the following appears:

LB	Theory and practice of education
[LB1025-LB1050.7]	Teaching (Principles and practice)
LB1045	Minor works
	Cf. LB41, Addresses, essays, lectures

leads to

LB Theory and practice in education
LB41 Addresses, essays, lectures. Pamphlets
For theory, methods, etc., see
LB1027, 1033, 1045

Other examples were somewhat more puzzling. Two other types of problems are worthy of mention. These are totally misleading captions and complex relationships.

In Subclass "LB Theory and Practice in Education", under "Child study", "LB1135 Adolescence" has a cf. note "HQ35, Sociology". In this instance the caption "Sociology" is very misleading. While subclass "HQ The Family. Marriage. Women" can be regarded as part of sociology, the general subclass for "Sociology" is HM, not HQ, and the caption at HQ35 is very precise - "Adolescents". Similarly, at "LB1075 Fatigue", the "Cf. LB3431, School hygiene" is given. It leads to "LB3431 Strain and overpressure. Mental capacity" which includes the reciprocal note Cf. LB1075, Fatigue." To be sure, [LB3401-LB3495] School hygiene" which is superordinate to "Strain and overpressure. Mental capacity" appears on the same page, but it is not the caption at the classification number referred to. If this is confusing on the printed page, it could be even more puzzling in the computerized system. Some other curious examples of this nature were found but without further time consuming investigation it was not possible to determine the precise magnitude of this problem in the LCC system as a whole. Some of these peculiarities appear to be errors, but in many cases they are a result of the way in which the cross-references have been structured.

In other cases the complexity of the relationships created among topics across the LCC system raises questions. In particular, puzzling examples were found in the use of cf. notes and "for ... see" references together in the same and related contexts.. While it is not the purpose of this study to judge the

intellectual quality of those relationships, two examples serve to point out complexities which may have implications for the future policies and editing procedures of the schedules after they are in machine-readable form. In the example below, two topics in two disciplines are related to the same topic, but not reciprocally to each other.

LB1137 Play, games, etc.
Cf. BF717, Psychology of play
GN454+, Games of primitive people
LB1177, Kindergarten
LB3031, Recreation in school management and discipline

HQ782 Play as a childhood activity
Including the social importance of play
Cf. BF717, Psychology of play
GF182.9, Guides for play
LB1137, Psychology of play for teachers.
Education through play

"Play" has a place in Education in Subclass LB "Theory and practice" and also in Subclass HQ "The family, Marriage, Woman." Both topics are related through a cf. note to BF717, Psychology of play". However, they are only related to each other through the cf. note which leads from HQ782 to LB1137 but not in the opposite direction. The caption in the cf. note "Cf. LB1137, Psychology of play for teachers. Education through play" seems to define the scope of the topic more clearly than the caption at LB1137 itself. Also when the classifier goes from HQ782 to LB1137, the note "Cf. BF717, Psychology of play" seems to suggest that "Psychology of play" is not at LB1137 but at BF717. The notes appear to be in conflict, and/or sending the user in circles. In addition this raises the question as to when cf. notes should be reciprocal and whether some cf. notes should only lead the user in one direction but be "traced" under the class number to which they refer. As already indicated, few reciprocal cf. notes were found in the sample and a majority of these notes only referred users in one direction. Also, there were numerous examples of variations in the use of cf. notes and "for ... see" notes as exemplified in Figures 17 and 18 below.

SUBCLASSES HM-HX

HQ

THE FAMILY. MARRIAGE. WOMAN

HQ

HQ503-HQ1064.U6 The family. Marriage. Home

HQ1060-HQ1064.U6 Aged. Gerontology (Social aspects). Retirement
Cf. BJ1691, Ethics...
HD7101+, Public medical care plans
HD7105+, Social insurance...
HG9426+, Annuities...
HV1597.5, Blind aged...
RA413.7A4 Private medical care plans
etc.**Figure 17. Cf. Notes**

The cf. notes in Figure 17 are in contrast to the "for ... see" references in Figure 18.

SUBCLASSES HM-HX

HV SOCIAL PATHOLOGY. SOCIAL AND PUBLIC WELFARE. CRIMINOLOGY HV

HV697-HV4959 Protection, assistance and relief

HV697-HV1493 Special classes

HV1450-HV1493 Aged.

Including medical advice and aid

Cf. GT 3100, Manners and customs

For blind aged, see HV1597.5For contributory old age pensions,
see HD7105+For developmentally disabled aged,
see HV3009.6.A35For private medical care plans,
see RA413.7.A4For private old age insurance, see
HG9426+For public medical care plans, see
HD7101+**Figure 18. "For ... See References**

Under "Aged" at HQ1060-HQ1064.U6 there are a total of 15 cf. notes in the Gale 1986 edition. In contrast, under "Aged: at HV1450-HV1493 there are

6 "for ... see" references and one cf. note. The lists are far from being identical, nor should they be expected to be identical. However, 5 of the 6 "for ... see" references lead to the same topics as the cf. notes. In 3 of the cases the captions match and in the other two they differ significantly. This problem in parallelism is not consistent across the LCC system, but there were a significant number of cases where this did happen. This raises two questions. Why were cf. notes used in one case and "for ... see" references in the other? How does the function of a cf. note differ from that of a "for ... see" reference? From Figure 18 it could be deduced that the note "Cf. GT3100, Manners and customs" leads to a related topic, while the "for ... see" references are used to connect a topic with other aspects of the same topic. However, this argument does not seem to apply to Figure 17 with respect to "Cf. BJ 1691, Ethics". This particular example could be an anomaly but similar examples of this kind of parallelism did not clarify the relationships. For every example of one kind of note there was an example of the other approach. As long as both "for ... see" and cf. notes are traced on the record for the classification number to which they refer, maintenance can be facilitated and users viewing the USMARC records for class numbers to which they are referred will benefit from tracings. However, there appears to be an intellectual problem which may deserve some attention in future.

7.4.3.2 "Prefer ..." Notes

"Prefer ..." notes are similar to cf. notes in some ways. They are attached to captions with valid class numbers, and to node labels representing spans of numbers. Occasionally they appear in the general notes at the beginning of a class or subclass. "Prefer ..." notes function as cross references and limit the scope of captions under which they appear, by recommending that certain aspects or treatments of subjects be classified elsewhere in accordance with the schedules. As the term "prefer" suggests, these note establish the "best" place for classifying documents on a particular topic. For example:

LB	Theory and practice of education
[LB3401-LB3495]	School hygiene
[LB3473-LB3479]	Feeding and clothing of school children
[LB3473-LB3475]	General works
LB3473	Documents
	Prefer LB3479 for individual countries

Here LB3473 is reserved for "documents" which are "General works" on the subject of "Feeding and clothing of school children", and the "prefer" note dictates that "Documents" on this subject with respect to a particular country are to be classified elsewhere in LB3479.

A total of 435 "prefer" notes occurred in the sample taken from the schedules. Most of these notes began with the word "prefer" but otherwise the format varied. With few exceptions, most "prefer" notes were brief and quite general in nature. Typical are such notes as: "Prefer individual country" and "Prefer classification with biography". Nevertheless some "prefer" notes carry references to single class numbers or number spans. While most "prefer" notes present no problems for an online system, where classification numbers are included they present the same problems as the classification numbers in references and cf. notes.

7.4.3.3 "Class here" and "Including notes"

The LCC system includes various kinds of notes which function as scope notes. Some of these are peculiar to the subject area, for example bibliographical history notes in the P schedules. Other scope notes are more general in nature and are applied generally across the schedules. The use of examples (without Cutter numbers attached) under such general captions as "Miscellaneous", "Other" and "Special" is one method used to at least partially explain the kind of topic which would be covered by such non-descript terms.

The most important and most prevalent of the general scope notes are the "Class here ..." and "Including ..." notes. "Class here ..." notes occur in both

schedules and tables in the LCC system under captions with valid class numbers, very frequently under node labels representing number spans and sometimes at the head of a class or subclass. Their primary function is to define the topics together with their subtopics. The format of "class here" notes varies, although there is some repetition of notes, for example "Class here general works only". In general "class here" notes are relatively brief. They rarely include classification numbers, but frequently are accompanied by "for ... see" references and/or cf. notes which function as "class elsewhere" notes and which aid in further definition of topics. Closely related are notes which use the terminology "class ... in" and "class ... with". These also function as "class elsewhere" notes.

In comparison with other notes and references "class here" notes are few in number in the LCC system. Only 474 such notes were present in the sample of 3934 pages. They were spread very unevenly across the sample. Most schedules had fewer than 10 such notes per 100 pages. However, H-HJ and HM-HX each revealed "class here" notes in 25% of the sample. Since "class here" notes do not carry classification numbers in most cases, they pose few technical problems for the online system. Any exceptions to this should be considered in the light of the problems identified with class numbers in other notes and references. One possible problem for the user of the online system is that scope notes often apply not only to the captions or node labels under which they appear, but they are also applicable to each topic subordinate to that caption or node label. For example in "HV Social Pathology. Social and Public Welfare. Criminology" at [HV7231-HV9960] Criminal justice administration" the following notes and references appear:

Class here comprehensive works on more than one aspect of justice administration, including prevention, detection, investigation and punishment of crime; and reformation of offenders;

For legal works on the criminal justice system, see K
Cf. HV6001+, Criminology

These notes and references apply to all of the subordinate topics in [HV7231-HV9960] but appear only once at the beginning of that section of the schedules on "Criminal justice administration." "Class here" notes under single specific topics with no subtopics present few problems. However, where these notes and references are attached to node labels they are applicable to each caption encompassed by the node label. In the printed schedules serendipity and easy flipping of pages aid the classifier. This will not be an option in the online system.

"Including ..." notes are scope notes of a different kind. They occur in both schedules and tables and are located under captions with valid class numbers and under node labels, as well as directly under some headings for classes and subclasses. Their primary function is to clarify and explain the contents of topics which might not otherwise be apparent from the captions alone. They extend topics and provide partial definitions of topics. Thus they differ from "Class here" notes which are intended to be comprehensive for topics under which they appear.

Nevertheless "class here" notes and "including" notes have some characteristics in common. With some exceptions, they tend to be brief, usually they do not carry classification numbers and they are frequently accompanied by "for ... see" references and/or cf. notes combining the functions of "class here" and "class elsewhere" under one caption. A few instances were found of "including" notes which were accompanied by "subarranged like" notes. Still other notes lacked the word "including" but the intent was the same, for example, as illustrated below, in "HQ The family. Marriage. Woman", all 4 types of notes are present under the caption "Children. Child development" at [HQ767.8-HQ792.2]:

Children. Child development.

Including the sociology of childhood;
Class here comprehensive works on child development
which emphasize social growth of the child;
For works on child development which stress mainly
the child's physical and psychological growth,
see RJ131
Cf. GT2450+, Customs and children
HQ783, Socialization
HV701+, Child welfare

How will the classifier know that these notes exist when consulting specific class numbers to which they apply? Such examples suggest the need for some means of displaying notes applicable to a whole class or large portions of a class in a way which makes them accessible to classifiers wherever they apply in the schedules.

Because "including" notes tend not to carry classification numbers they will present few problems in converting to an online system. However, should any classification numbers be identified in such notes they will need careful checking as it is reasonable to assume that these classification numbers could present the same problems as those in other notes and references.

7.4.4 RECOMMENDATIONS

The foregoing examples indicate a number of concerns. It is recommended that:

- a) Captions be included for all class numbers referred to in references and notes, and that the captions be correct for the class numbers to which they refer;
- b) Provision be made for classifiers to view simultaneously the class number and caption referred from and the class number and caption

referred to, and for the display of both in their hierarchical contexts;

- c) All notes and references carrying class numbers be checked and edited as necessary to ensure that the class numbers and captions match;
- d) Consideration be given to the problem of general notes and references which apply to a whole class or subclass, or large portions thereof to determine how the classifier can be kept aware of these notes and references in the online system; and
- e) In the light of the amount of editing of notes and references which appears to be required in the conversion, where appropriate this editing take place after the schedule data is entered into the computerized system. The ability to search class numbers and captions across schedules should greatly facilitate the editing process.

7.5 Footnotes

Many pages in the schedules contain footnotes. Most of these footnotes serve as instructions to classifiers on the manipulation of the schedules. They provide directions which locate tables by table number and/or page number, or both; as well they give directions on how to create classification numbers using tables, or they instruct classifiers that particular topics are "divided like" or "subarranged like" other sequences of classification numbers. Four types of footnotes with minor variations frequently found in schedules were the following"

"For ... see ..."
"Add number in table to ..."
"Subarranged like (by) ..."
"Divided like ... "

A few footnotes are unique, or unique to a particular schedule. Also occasionally small tables or subarrangements for successive Cutter numbers were found in footnotes. For example in Class C beginning at CR5010 a

footnote indicates to classifiers that "One number countries [are] subdivided as follows: .A2A-Z, General works; .A3-Z, Particular orders, A-Z". Obviously the most frequently used footnotes parallel in wording notes attached to individual captions in the schedules. However, the function of notes and footnotes is often different. Notes provide additional information about the content of the captions under which they are displayed, whereas footnotes are usually instructional in nature. Examples of some of the most frequently used footnotes are as follows:

From B-BJ:

"For subarrangement by language, see Table 6, p. 227."

"For subarrangement under philosophers, see Tables, pp. 132-136."

From HM-HX

"Subarranged like HS1510." [leads to a table]

"For Table V, see pp. 105-204. Add country number in Table to 4760."

"For subarrangement, see p. 84." [leads to tables]

Some footnotes are linked to generic captions as is the case in some of the examples above where one footnote in B-BJ is linked to the caption "Individual philosophers" rather than to the name of each philosopher; and another in HM-HX is linked to "By country or region" rather than to the name of each country. Some footnotes lead to more than one table. In such cases only the page number is given and the specific table is not specified. Other footnotes provide for a choice to be made depending on conditions present as exemplified in the following:

"For Table V, see pp. 195-204. Add country number to 4040 or 4190 as the case requires".

"For subarrangement under countries, unless otherwise provided for, see Table I-II, p. 133."

"Otherwise provided for" signifies that this footnote applies to each country which does not have a Table number attached to its caption. For example, France (I); Germany (I, modified) are to be handled by Tables I and

II (e.g. CS93 Greenland; CS1311 Kuwait; CS 1647 Zaire).

In the case of the four most frequently used types of footnotes 756 footnotes were located in 3934, or 19.2% of the pages in the sample. In most cases there was not more than two footnotes per page. However, the same footnote could be repeated over many pages representing many records. The earlier example "Individual philosophers" in B-BJ is a case in point. Based on the average of 33 captions per page, editing and transferring the footnote data to 10 pages of individual philosophers could involve the editing of 330 records. This kind of editing will be required wherever instructions such as these apply to individual captions but are linked only to the generic caption in the printed schedule.

The footnotes are unevenly distributed across the schedules. The average number of footnotes in schedules such as B-BJ and C is one footnote for every 2 pages in the sample, whereas, not unexpectedly, schedules K and P have almost no footnotes. This is because of the peculiarities of structure of the K schedules where the schedules are small and the tables take up the major portion of each subclass. Also Class P is organized with a separate schedule for tables (P-PZ) so names of authors are linked directly to table numbers and footnotes are not required.

With respect to the conversion of the schedules to machine-readable form, considerable editing will be required. All of the footnotes are page-oriented and will need to be eliminated and absorbed into the records for the captions to which they apply. Each specific caption will need to be linked to its table where tables are involved. "Divided like" and "Subarranged like" footnotes would either become notes under individual captions or be absorbed by enumerating in the schedules the class numbers covered by these instructions. This will depend on the policy adopted in dealing with "divided likes". The same approach could be used for both the printed and machine-readable products. However, the elimination of "divided like" will increase the size of

some printed schedules.

7.5.1 RECOMMENDATIONS

It is recommended that:

- a) Footnotes be eliminated and the information in them be edited and absorbed into the schedules for both printed and online versions of LCC;
- b) Care be taken that all footnotes attached to generic superordinate captions be linked directly to the individual captions to which they apply;
- c) Consideration be given as to how footnote data will be handled in the printed schedules in future;
- d) Small tables or successive Cutters found in footnotes be treated in the same way as other small tables throughout the schedules; and
- e) "Divided like" and "Subarranged like" footnotes be handled in the same way as other occurrences of these devices throughout the schedules.

7.6 Class Numbers

Statistics derived under the topic "class numbers" ("Schedule Sampling Form", Appendix B) were collected to provide data on the approximate number of records which would need to be converted to machine-readable form. Three kinds of data were collected on classification numbers which actually appeared on the pages in the sample. These were the number of single classification numbers (including Cutter numbers appearing in lists in the schedules), the number of spans of numbers, and the number of independent "see" references. In the 3,934 pages in the sample there were 102,307 single class numbers, 5,229 number spans and 9,675 independent "see" references, for a total of 117,211 records. However, an additional 15,002 captions were located in the analysis of the hierarchical levels which function as node labels and therefore had no visible classification numbers attached. Combining these figures, it was

found that there were 132,213 records in the sample (i.e. 33.6 records per page). Since a total of 12,267 pages were contained in the schedules and tables examined, it is estimated that these particular schedules could possibly contain approximately 412,171 records. However, this figure must be considered to be "very" approximate for several reasons. The 412,171 records do not include records which would be added if the "divided like" instructions were eliminated and those records added to the schedules. If such were the case, it is estimated that 10,098 additional records would be added to the records in the sample for a total of approximately 142,311 records. Applying these data to the entire schedules, the potential exists for 443,496 records. Additionally, new classification numbers will have been added to the schedules since the data in this project were collected and it must be remembered that some tables are combined with common sets of captions. This figure would be considerably increase the number of records in the schedules depends on how these combined tables are handled. Taking these unknowns into consideration, it is reasonable to assume that the conversion will involve more than 500,000 individual records.

7.7 *Divided Like Instructions*

The "divided like" mechanism is a device which instructs classifiers to arrange the topics covered by a particular span of class numbers in the same way as the topics contained in another span of numbers where the topics and subtopics have the same characteristics. That is, one display of topics and subtopics provides the pattern for one or more other topics with the same subtopics.

"Divided like" devices are somewhat analogous to tables because they perform a similar function in a different way. Both are devices which minimize the bulk of the printed schedules and would not necessarily have to be retained in a machine-readable version of LCC. While the elimination of tables presents many problems, "divided like" instructions can be eliminated in

most cases by simply enumerating the topics and their class numbers in the schedules.

The instructions for dividing like are provided for in several ways. The instructions are variously called "divided like", "subdivided like" and "subarranged like" and are located in notes directly under the caption to which they apply or are footnotes on pages of the schedules where the instruction applies to several spans of numbers. "Divided [etc.] like" instructions should not be confused with notes which instruct the classifier that a number is "subarranged by" or "subdivided by". The word "like" signifies that the pattern of arrangement of topics in one place in the schedules is to be used in another, whereas the word "by" indicates that class number is to be extended for additional topics and subtopics, as for example in the use of Cutter numbers. Some "divided like" instructions lead to spans of class numbers elsewhere in the schedules, others lead to tables. Divided [etc.] like, instructions do not usually apply to the division of subarrangements (or Cutter lists), but could apply if a Cutter number sequence were "divided like" another Cutter number sequence.

One of the purposes of the study was to determine the impact of eliminating the "divided like" instructions from the schedules and replacing them with the actual captions and class numbers. In the statistical count 1,306 occurrences of "divided like" instructions occurred in the sample. These instructions encompassed 10,098 captions with an estimate of a possible 31,325 records which would need to be added to the schedules if the "divided likes", exclusive of those which lead to tables were eliminated. This would increase the number of captions by 7.6%.

Since "divided like" instructions are space saving devices, it seems logical to eliminate them from the machine-readable version of the schedules. Their elimination would have the added advantage of making the schedules easier to use. The number of records which would need to be added is not excessive in

terms of the magnitude of the whole conversion project. Whether they should also be eliminated from the printed version would need to be considered.

There were 132,171 captions including "see" references in the sample with a potential for 412,266 records. The addition of a possible 31,325 records, or an increase of approximately 7.6% does not seem excessive in this case.

7.7.1 RECOMMENDATIONS

It is recommended that:

- a) Where "divided like" instructions lead to spans of numbers for which the class numbers and captions are enumerated in the schedules, the class numbers together with their captions be enumerated in the schedules and the instructions be eliminated;
- b) Consideration be given to the advisability of this same elimination from the printed schedules.

7.8 Caption and Record Size

At the outset of this study, it was thought useful to collect data that might be used in determining record size. To this end, data were collected on caption size (Appendix B, no. 8). However, captions are only one of numerous types of data elements which will appear in the machine-readable classification records. These data alone would be an insufficient basis for estimating record size. Nevertheless, there are some general observations that can be made on three types of data elements from the study that might be helpful in decision making. These elements are classification notation, captions and notes.

7.8.1 Classification Notation

With reference to Chan (1990) and the analysis of the schedules, classification notations, as they appear in the schedules, may consist of two or

more of the following:

- a) One, two, or three capital letters
- b) Whole numbers, 1-9999
- c) Decimal extensions
- d) First Cutter number
- e) Second Cutter number; and
- f) Date of publication

Examples of the first five data elements were located in the analysis of the schedules. No examples included dates, since these apply to specific publications. Decimal extensions with up to 3 digits and Cutter numbers with up to 4 digits were located in the schedules. Based on these data it can be estimated that some classification notations could be upwards of 25 characters in length. Individual class numbers will be somewhat shorter, but when the notations for topics and node labels are made explicit in the records the number of characters doubles. Some examples of notations (expanded to make all letters and digits explicit) found in the schedules are the following:

DS748.164

HD[8039].C6523

HD[9259].A95-[HD9259].A954

KJV1881-[KJV]1890

PA4333.A1A-[PA4333].Z

PT1670-[PT]1678

7.8.2 *Captions*

For purposes of this study, the number of characters in all captions, including node labels and independent see references were counted. A template was used to determine the number of characters, including blank spaces. Statements, such as "A-Z" "By date" and "(Table I)" were categorized as instructions, rather than as caption data, and were excluded when calculating the number of characters.

In recording the data, captions were grouped by size using increments of 5 characters. For example, captions fell into; up to 5 characters; 5 to 10 characters; 10 to 15 characters, etc. up to 120 characters and beyond. The largest number of captions in the sample (31,654 or 23.9%) fell into the 10 to 15 character range, followed by the 15 to 20 character range (26,189 captions) and the 20 to 25 character range (17,076 captions). Together these 3 ranges contained 56.6% of the captions in the sample. This figure was heavily supported by the caption "General works" which occupies 13 character spaces and appears many times in the LCC schedules. There were 402 captions in the sample which exceeded 120 characters. These were spread unevenly across the schedules but the majority were located in Subclasses BR-BV. Seventy-two percent of the total captions had from 1 to 40 characters and 78% had 1 to 65 characters. See references are sometimes longer than other captions, but some ordinary captions are also very long. For example, the Subclass heading for HV "Social Pathology. Social and Public Welfare. Criminology" requires 58 character spaces if two spaces are left after each full stop. If a change in policy were to include captions following the class numbers referred to in the see references, this would change the statistics on the length of the captions.

7.8.3 *Notes and "For ... see ..." References*

Notes and "for... see ..." references vary greatly in length depending of the nature of the data. They tend to be more variable than either classification notations or captions. In the course of the analysis, some notes were found that were more than 325 characters long.

"For... see..." references contain either the words "For... see..." or "For works on... see ...," together with a caption, or caption-like statement, and a classification number (e.g. "For local societies and associations, see HV98+). Many such references are short, but some are quite lengthy. For example at HV7575-HV7625 in the section on "Criminal Justice Administration" the

following note appears"

For reports of departments of public safety,
including fire, police, and in some cases,
other departments see JS13, unless some special
reason appears to the contrary.

Some notes, in particular cf. notes, are somewhat more predictable and have a fairly regular format. Cf. notes consist of the abbreviation cf., a classification number and, in most cases, a caption. The classification number may include a Cutter number but number spans are rarely enumerated, the plus (+) sign being used instead. However, there is an exception for every rule. Under "Ancient Italy. Rome to 476" at "DG27-DG41 Geography. Description and travel" the note "Cf. G84-88, History of ancient geography" appears. "Prefer", "Including" and "Class here" notes also vary considerably in length.

The number of notes attached to a particular record will also affect record size. The sample located 23,242 notes and "for ... see..." references of some kind linked to 122,538 captions (exclusive of independent see references). As a result it appears that approximately 1 caption in 5 has a note or reference attached, suggesting that the impact could be small. In actual fact these notes and references are spread very unevenly across the schedules. For instance a single topic may have 10 or 12, or more cf. notes, while many other headings may have none. The clustering of notes or lack of them often depends on the nature of the subject area.

7.8.4 Other Classification Data Elements

The USMARC record format for classification provides for numerous other data elements, some of which are non-schedule elements. However, there are two other data types from the schedule elements which should be noted. These are the tracings for the references, and the index terms. For all references and all notes which function as references, tracings are to be added to the machine-readable records for the topics referred to, increasing the note-like data

added to the system. Also there is a place in the record for index terms from the indexes which are part of most individual schedules. Without further study there is no way of knowing what the average number of index terms there are to be added to each record, for the present the index terms and their impact on the records is an unknown.

While this discussion of the analysis cannot provide an accurate picture of record size required to accommodate the greatest possible amount of data, it may at least identify some things which need to be considered in designing and adjusting the proposed machine-readable record.

8. SUMMARY OF RECOMMENDATIONS

Specific recommendations made in this report have been set out at the end of each section of the report with which they are concerned. The characteristics of LCC which could present problems for its conversion to machine-readable form have been pointed out. Suggestions and recommendations for solving some of these problems have been presented. The individual recommendations are not discussed in detail here; rather they have been categorized and summarized in a brief overview.

Most important to the conversion is the fact that the environment of the printed pages will no longer exist. All data must be made explicit and relationships among data must be correct and understandable. Incomplete classification numbers must be complete and the hierarchical relationships carefully maintained; links between schedule data and tables must be precise. Footnotes which are intended to serve for whole pages of captions will need to be eliminated and the information transferred to the individual captions to which they apply. Screen displays must be intelligible and efficient for the editors of LCC, classifiers and other users.

Insofar as possible, efforts should be made not to allow the printed format

to hinder innovation. It is important to make the best possible use of new technology and where feasible to improve on schedule display and manipulation of the printed schedules. Specifically, serious consideration should be given to the elimination of "divided like" instructions and to the development of an expert system approach to the manipulation of the tables. Also, there are several situations where the editors of LCC and the classifiers need to simultaneously consult various parts of the schedules. Examples of situations are those where tables are being applied and where the use of references and notes require users to consult two or more locations in the schedules at the same time. Current technology provides for such devices as windows, split screens and notepads which if incorporated into the system to advantage could greatly improve efficiency in the use of the schedules.

The most obvious factor in the conversion is the tremendous amount of editing which will be required. For ease and speed of conversion, decisions need to be made on when and how the editing should take place. If the schedules are keyboarded and editing is incorporated into the keyboarding, progress will be extremely slow. With the use of the MARC format computerized scanning may not be an option. However, the creation of an LCC database using a scanning operation and then coding the scanned text and transferring it to the USMARC format should be investigated. There are some data, particularly notes and references, which could probably be edited more effectively online. In the light of this, the feasibility of computerized scanning of the schedules should be given serious consideration. Recent experience with the Bliss Bibliographic Classification schedules has shown that classification data can be scanned in the conversion to machine-readable form, with as much as 95% accuracy. The major problem is that the hierarchies are often disturbed. However, it is more efficient to correct this online than to key it in in the first place. In LCC, there is some editing, such as corrections in the reference and note data, which would be most efficiently done after the data are in machine-readable form and searching can be done by computer.

Scanning might also be considered in conjunction with the indexes to the individual schedules. It would be useful to have these indexes in machine-readable form during the conversion process. If both the terms and the classification numbers could be searched by computer, it would be possible to quickly locate all index terms which lead to specific classification records for incorporation into those records.

Ultimately, it will be those designing the system and editing the input who will make the final decisions on what editing is done and how it can best be accomplished.

9. FURTHER RESEARCH

This is only the first step in achieving four long term goals which could eventually see LCC being used by the editors of the LCC system, by classifiers at work stations with all of the cataloguing tools online, and by users as a search tool in online catalogues. The USMARC Format for classification is now in place and already being revised and extended during the testing stage. This process will undoubtedly continue into the future. Nevertheless, if an ideal system is to be achieved, there are further studies needed in a number of areas.

First of all, this report does not include an analysis of the indexes to the individual schedules. However, it is known that these have been developed on an individual basis in conjunction with specific schedules. They have been created at different levels of analysis and without reference to any overall guidelines for compiling indexes to LCC. Nevertheless, they are rich in vocabulary and should not be overlooked. A comprehensive study of the indexes is needed to determine problems that the indexes present in terms of terminology, structure and depth of analysis. While specific disciplines and subdisciplines may have special indexing requirements, it should be possible to develop some general guidelines for indexing the schedules, incorporating

special requirements where necessary. These guidelines could then be the basis for future editing and development of the indexing vocabulary.

If it is assumed that for the foreseeable future there will need to be a printed as well as a computerized version of LCC, several questions on the differences between the two versions need to be addressed. Will the contents of the schedules be identical? Two examples come to mind. If the "divided like" instructions are eliminated from the computerized LCC, will they also be eliminated from the printed LCC? Keeping in mind the current page format in the printed product, will all of the schedule data made explicit for the computerized version be displayed in the printed version? Would such a decision make the printed pages too cluttered? Confusing to users? Decisions made on this issue would surely have implications for MARC coding. How will the indexes in the printed version be affected by the use of the machine-readable data? Will there be indexes in individual volumes? One overall index? Both? How will the structure of the printed indexes be affected? The production of printed indexes to classification schemes using the computer has, up to now, not been very successful. The published index to the 1st International Medium Edition of the Universal Decimal Classification illustrates many of the problems. With respect to tables, the location of internal tables and subarrangements in close proximity to the schedules to which they apply and their frequent duplication, or near-duplication, is convenient for the users of the printed schedules. If there is a consolidation of tables, and the elimination of some tables in the computerized version, how will it effect the printed product? These are but some of the questions that must be addressed with the respect to the relationship between the two types of products and their output from a common set of machine-readable records.

For the immediate future, this report recommends that research be carried out to investigate ways and means of improving the process of manipulating the tables and Cutter subarrangements. It seems probable that an expert system, or systems, could be developed which would permit the mechanical

manipulation of tables to be carried out by computer, leaving the intellectual decisions to the classifiers. For the more distant future, research should be undertaken to determine the feasibility of eliminating the tables (with the possible exception of Cutter tables) altogether. Tables are the equivalent of "divided like". They provide for repetitive division of sections of the classification schedules which have topics in common and which could be inserted into the schedules. The amount of conversion involved could be massive and might not be suitable for printed schedules, but it is a possibility which should be considered for the future when there may no longer be a printed product.

Finally, if eventually, the schedules are to be made available in the online catalogue, research needs to be carried out to determine what schedule data will be displayed to the user and how it will be displayed. Catalogue users do not want to see tables and they do not want to see instructions which are intended exclusively for classifiers. However, what has to be added to the schedule display to help the users of an online catalogue version of LCC? Such questions as: What kind of user-machine interface is needed? What kinds of displays are needed? What schedule data should be suppressed in the OPAC display version? Excellent ground work in this area has already been laid in the development of the Electronic Dewey. Many of the same techniques may be applicable to LCC, but it has unique characteristics which would need to be investigated.

There are surely other areas for further research which have not been mentioned here but which will be revealed as the computerized system develops and it is abundantly clear that there are enough projects to keep researchers interested in classification busy through the 1990s and beyond.

10. CONCLUSIONS

When this study was initiated, there was little certainty that the conversion of the LCC schedules to machine-readable form was even feasible. Five years later it appears not only to be feasible, but probable. A format has been developed and is in the testing stage and problems, while numerous, appear to be surmountable. Nevertheless the study has raised questions yet to be answered as well as suggested solutions.

Can the LCC schedule data be stored and manipulated so that these can be accessed intelligently, usefully displayed and efficiently manipulated for effective use in a computerized environment? While only the results of testing the USMARC format for classification can determine the precise answer to this question, the investigation suggests that the answer is "yes" if care is taken to determine exactly how users will wish to access the data, and if care is taken to ensure that the capabilities of the technology are used innovatively and to the fullest. The process of conversion provides an opportunity for the development of new ways of handling and using classification data.

What will conversion to machine-readable form entail and how should it be accomplished? The content analysis makes it very clear that there is a great deal of editing to be done in the process of conversion. Among the essential requirements are the following:

- a) All data which is only implicit in the printed volumes must be made explicit in the machine-readable version;

- b) Relationships among captions, references, notes, class numbers which are obvious and easily understood on the printed page but are not necessarily precise in machine-readable terms must be clarified for computer input. Specifically, hierarchical relationships must be precise, so that data is not "lost" in the computer system.

Relationships which exist in the LCC system through the use of references and notes and links between schedules and tables must be clear to the schedule editors, classifiers and other users. In short, the "black box" effect must be minimized.

- c) Insofar as possible, procedures and devices required for the use in the manipulation of the printed volumes but not needed in a computerized system, should be adjusted to the machine environment. Specifically, duplication of tables should be reduced to a minimum; machine-aided manipulation of the tables should be incorporated into the system where possible and "divided like" instructions should be eliminated by inclusion of the appropriate class numbers and captions in the schedules.

What kind of editing and how much will be required as part of the conversion? As has already been stated, there will be much editing? There are at least two kinds? The simplest kind of editing is the simple completion of data which is not explicit in the printed schedules (for example the completion of classification numbers), the provision of spans of classification numbers to accompany node labels, and the addition of classification numbers to Cutter numbers which are part of the schedules. The second kind of editing is more complex and includes the realignment of hierarchies where the alignment on the printed page is not precise or lists of names have interspersed node labels which could affect the display. Other examples of this latter type of editing could include the insertion of class numbers and captions which are covered by "divided like" and "subarranged like", the clarification of notes and references to ensure that they lead to the proper place in the schedules and the elimination of footnotes and the linking of those instructions to specific captions to which they pertain.

What improvements could be incorporated into the system, so that LCC could be a better managed classification than it is in the printed mode? It is

assumed that the implementation of some of the recommendations in this report will lead to a computerized version of LCC that will include screen displays and sophisticated search capabilities which will permit better access to classification data, instructions which will lead to fewer errors in the data, and an improved environment for ongoing editing and revision of the schedules and more effective use by classifiers, including the more efficient manipulation of both schedules and tables.

What should the long term goals be? These are the goals which were established at the outset of the project - the development of an online interactive editorial support system, the provision of an online version of the schedules for the use of classifiers, the production of LCC in a variety of physical forms, including printed volumes, MARC tapes and CD-ROM products, and the use of the classification as a search aid in online catalogues.

A final question posed in the introduction to this report was "What would be the sheer volume of the work to be accomplished?" In general, the volume is enormous, but the data gathered suggests that it is not an insurmountable task, given the time, effort and resources to accomplish it. Throughout the report the statistical data gathered were examined in conjunction with specific characteristics of the schedules and have been used to support individual recommendations. As well, separate from this report, the data have been used to create brief profiles of each schedule in order to provide the editors with a brief overview of the schedules and their idiosyncrasies which may affect the editing process.

APPENDICES

APPENDIX A

Schedules Used in the Study

To make it possible to sample the most complete in-one-volume versions of the LCC schedules available at the time, the majority of schedules used in the study were the Library of Congress Classification Schedules Combined with Additions and Changes through 1986. These were graciously supplied by the Gale Research Company. However, a complete set of the 1986 edition was not available at the beginning of the study and the Gale schedules were supplemented by the latest editions of the LCC schedules for G, H-HJ, R and Z. The following were the schedules used:

A	General Works
B-BJ	Philosophy. Psychology. Religion: Philosophy, Psychology
BL-BQ	Philosophy. Psychology. Religion: Religions, Hinduism, Judaism, Islam, Buddhism
BR-BV	Religion: Christianity, Bible
BX	Religion: Christian Denominations
C	Auxiliary Sciences of History
D	History: General and Old World
DJK-DK	History of Eastern Europe (General). Soviet Union. Poland.
DS	History of Asia
E-F	History: America
G	Geography. Maps. Anthropology. Recreation. (Library of Congress. 4th ed. 1976)
H-HJ	Social Sciences. Economics (Library of Congress. 4th ed. 1981)
HM-HX	Social Sciences: Sociology

J	Political Science
K	Law (General)
KD	Law of the United Kingdom and Ireland
KDZ, KG-KH	Law of the Americans, Latin America and the West Indies
KE	Law of Canada
KF	Law of the United States
KJV-KJW	Law of France
KK-KKC	Law of Germany
L	Education
M	Music and Books on Music
N	Fine Arts
P-PA	Language and Literature: Philology, Linguistics, Classical Philology, Classical Literature
PA	Language and Literature: Byzantine and Modern Greek Literature; Medieval and Modern Literature
PB-PH	Language and Literature: Language and Literature: Modern European Languages
PG	Language and Literature: Russian Literature
PJ-PM	Language and Literature: Languages and Literatures of Asia, Africa, Oceania, America, Mixed Languages; Artificial Languages
PN,PR,PS,PZ	Language and Literature: Literature (General); English and American Literatures; Fiction in English; Juvenile Belles Lettres
PQ, Part 1	Language and Literature: French Literature

PQ, Part 2	Language and Literature: Italian, Spanish, Portuguese Literature
PT, Part 1	Language and Literature: German Literature
PT, Part 2	Language and Literature: Dutch and Scandinavian Literature
P-PZ	Language and Literature: Language and Literature Tables
Q	Science
R	Medicine (Library of Congress. 5th ed. 1986)
S	Agriculture
T	Technology
U	Military Science
V	Naval Science
Z	Bibliography and Library Science (Library of Congress. 5th ed. 1980)

Appendix B

SCHEDULE SAMPLING FORM

Schedule: _____ Sample No. _____ of _____ Page No. _____

Circle one: Gale LC Date: _____ Total No. of pages in schedule: _____

1. Hierarchical Levels. Circle the no. of captions at each indentation level.

Level	No. of Captions																			
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
7	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
8	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
9	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
10	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
11	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
12	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
13	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
14	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
16	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
17	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
18	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
19	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

2. Internal Tables and Subarrangements.

Number on page: tables _____ subarrangements _____

Circle response; write no. within circle if more than one of a type on page.

Table types are: topical/form geographical both

Table level is: first level second level third level

Table arranges: numbers Cutters both

Schedule: _____ Sample No. _____ of _____ Page No. _____

3. Notes and References.

Circle type; write no. within circle if more than one of a type on page.

Reference: "... <u>see</u> "	Notes: "Cf. ..."	Divide/Subarr. like _____
	"Including ..."	Subarrange note _____
	"Class here ..."	Subarrangement of
	"For ... <u>see</u> ..."	captions _____
	"Prefer ..."	

Other type: _____

4. Footnotes.

Circle type; write no. within circle if more than one of a type on page.

"For ... <u>see</u> ..."	"Subarranged like ..."
"Add number in table to ..."	"Divide like ..."

Other _____

6. Number of numbers. On the sampled page, count:

Single numbers _____ Number spans _____ See references _____

7. Divide like numbers. On the sampled page, count:

Occurrences of "divide like" _____

Number of class numbers covered by each "divide like"

8. Caption size.

Using a template, count the number of captions with the number of letters falling in the indicated range:

5	10	15	20	25	30	35	40
_____	_____	_____	_____	_____	_____	_____	_____
45	50	55	60	65	70	75	80
_____	_____	_____	_____	_____	_____	_____	_____

APPENDIX C

Instructions for Sampling the LCC Schedules

The following instructions were used in the sampling of the 3934 pages of schedules and tables using the form found in Appendix B.

Sample size:

100 pages of schedules and tables (for volumes of more than 100 pages). 50 pages of schedules and tables (for volumes with fewer than 100 pages). For volumes with fewer than 50 pages, sample all pages.

Random selection:

Select pages to sample by flipping through a volume. The pages need not be selected in consecutive order from front to back. However, do not sample a page more than once. If there are sufficient pages to choose from, try to avoid sampling long consecutive runs of pages.

Sampling form:

The form consists of two pages. Use one form for each page selected. The following information about the schedule and table data should be entered at the top of the first page of each form:

1. Schedule:	Class and subclass letters covered by the volume;
2. Date:	Publication date of the volume.
3. Gale/LC:	Publisher of the volume.
4. Sample no.:	Number of the sample and the total number of samples taken from that volume; e.g. "Sample no. 10 of 100".
5. Page no.:	Number of the sample page in the printed volume;
6. Total No. of pages:	Total number of pages in volume including tables and excluding a synopsis, outline, index and appendices (if any).

The numbered sections on the sampling form correspond to the questions on the original profile document. Annotate the form as needed if sufficient blanks or numbers are not provided and to take care of situations where data appear to be irregular, unusual or unique and will be of importance to the principal investigator in the data analysis.

The sampling instructions below follow the order of data categories on the sampling form.

1. Hierarchical levels:

Count the number of captions at each indentation level on the sample page. Do not count the caption at the top of the page which ends with the word " - Continued" or any caption above the "continued" caption.

The caption of the class/subclass letter(s) which is the page header is the level 1. Count this caption only if the sample page is the first page of the class/subclass.

Count captions with bracketed numbers which are also see references in both sections 1 and 3 on the sampling form.

Captions in tables which are farthest to the left are at level 1.

Centred headings which are in the centre of the page and subordinate to the page header are counted as being at level 2 (in some exceptional cases level 3 or 4). The captions which are subordinate to these centred headings begin at 1 level lower than the centred heading immediately above, even though visually they may appear to be at the same level as higher level captions which are subordinate to the page header but not subordinate to the page header and the centred heading(s).

Included in the count are: individual captions, see references, examples which begin with a class number of the particular class being sampled, or Cutter numbers and Cutter lists which are an integral part of the schedule at that point. Do not count as captions for purposes of hierarchical levels, examples from other parts of the schedules.

Where there are lists of names or topical subjects with node labels interspersed throughout the alphabetical listing and these labels cause inconsistencies in the levels at which the topics or names appear, count the captions as if the node labels were not present (e.g."Individual philosophers" in B-BJ).

2. Internal tables and subarrangements:

Count the number of internal tables and subarrangements appearing on a sample page. Indicate the total number of each. In indicating the type, level, etc. write the number within the circle if there are more than one of that kind on the page. For example, if there are two geographical tables write a "2" inside the circle around "geographical".

3. Notes and references:

Count each note and reference on a page by type and record the total number of each type. If there is more than one note of the same type under the same caption each counts as a separate note. If two notes are combined in one sentence count them as separate notes. For example a "class here" note combined with a "for ... see ..." note in the same sentence counts as two notes, each one of a different type. Make a note of any types of notes and references different from those listed on the form. For example identify "e.g." notes as scope notes. Write brief notes on any peculiarities found in the notes.

4. Footnotes:

Apply the instructions as for "Notes and references" above. Identify separately and count notes which are "instructional" notes as opposed to notes which contain data.

5. Question 5 on the original profile was dropped.

6. Number of numbers:

Count the number of single numbers, number spans and see references on a sample page.

7. Divided like numbers:

Count the number of times the "divided like" technique is used on the sample page. Refer to the span of class numbers to be divided like, and record on the blanks the number of class numbers encompassed in each occurrence of the "divided like". Identify any "divided like" instructions which lead to tables, but do not count the captions in the tables.

8. Caption size:

Make a template for the appropriate print type size (Gale and LC schedules will differ) with marks at intervals of 5 characters. Run the template down the sample page, and record the number of captions with the number of letters falling within the indicated range. For example, if ten captions have a total number of letters between 15 and 20, record "10" on the blank below "20". Note: in multi-word captions a blank space between the words counts as a character.

APPENDIX D

REFERENCES

Beall, Julianne. 1992. "Editing the Dewey Decimal Classification Online: The Evolution of the DDC Database". Classification Research for Knowledge Representation and Organization: Proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, June 24-28, 1991. Edited by Nancy J. Williamson and Michèle Hudon. Amsterdam: Elsevier. (FID 698). Pp. 29-37.

Chan, Lois Mai. 1986. "Library of Congress Classification as an Online Retrieval Tool: Potential and Limitations." Information Technology and Libraries 5: 181-192.

Chan, Lois Mai. 1990. Immroth's Guide to the Library of Congress Classification. 4th ed. Englewood, Colo.: Libraries Unlimited.

Cochrane, Pauline Atherton and Karen Markey. 1985. "Preparing for the Use of Classification in Online Cataloging Systems and in Online Catalogs". Information Technology and Libraries 4: 91-111.

Freeman, Robert R. and Pauline Atherton. 1968. "AUDACIOUS: An Experiment with an Online Interactive Reference Retrieval System Using the Universal Decimal Classification as the Indexing Language for the Field of Nuclear Science. Information Transfer: Proceedings of the American Society for Information Science. Annual Meeting, Columbus, Ohio, October 20-24, 1968. New York: Greenwood Press. Pp. 193-199.

Library of Congress. Network Development and MARC Standards Office. 1990- USMARC Format for Classification Data: Including Guidelines for Content Designation. Washington, D.C.: Library of Congress, Cataloging Distribution Service.

Markey, Karen and Anh Demeyer. 1986. Dewey Decimal Classification Online Project. Dublin, Ohio: OCLC Online Computer Library Center.

Rigby, Malcolm. 1981. Automation and the UDC, 1948-1980. 2nd ed. The Hague, Netherlands: International Federation for Documentation.

Rigby, Malcolm. 1974. Computers and the UDC: A Decade of Progress, 1965-73. The Hague, Netherlands: International Federation for Documentation.



UC-952-320